

Struggling with the flu

The shortages of flu vaccine in the United States this autumn have laid bare some troubling weaknesses in the nation's public-health system.

This month, the United States has been experiencing a public-health fiasco about which government officials had been repeatedly forewarned.

Since the anthrax attacks of 2001, health experts have been highlighting the weak condition of the nation's system for purchasing and distributing vaccines against potential bioterror agents. But this autumn's shortage of flu vaccine — triggered when British regulators found contamination in vaccine from a plant in Liverpool that was supposed to supply half of it — graphically demonstrates their point.

The flu vaccine is tricky to produce, because it must be grown in chicken eggs months in advance. But that is not the root of the problem. As an Institute of Medicine report expressed it last year: "The public-private partnership that has formed the foundation for purchasing and distributing vaccines over the past 50 years is showing signs of erosion."

The United States relies on private companies to make flu vaccine, many of which have decided that they could better serve their stockholders by making blockbuster drugs that boost sex drive or cut high blood pressure. Unlike Canada or Britain, the United States also relies on the private sector to distribute flu vaccines at grocery stores, pharmacies and doctors' surgeries.

As a result, according to an official from the Government Accountability Office who testified at Congress last month: "There is no system in place to ensure that seniors and others at high risk for complications receive flu vaccinations first when vaccine is in short supply." This was demonstrated all too vividly last week, when a 79-year-old Californian woman stumbled and died after having to

queue for four hours for the vaccine in a grocery-store parking lot.

President Bush has been trying to shift the blame on to trial lawyers, claiming on the campaign trail that excessive legal liability has deterred drug companies from making vaccines. But that isn't why they have stepped away from producing vaccines; the real reasons are low profit margins and a lack of reliable demand.

In July, Bush signed the Project Bioshield Act, which is supposed to lift the obstacles that prevent drug companies from developing the medicines needed to defend against biological attack. But it isn't clear that the Department of Homeland Security, which sets the priorities for Project Bioshield, regards flu as falling under its domain.

The Department of Health and Human Services, under the leadership of health secretary Tommy Thompson and the supervision of Congress, has failed to take steps to ensure a steady demand for flu vaccine, or guarantee a price for it. It could have responded to previous warnings and issued an estimate for the number of vaccines that the nation would need in the coming year, while promising to purchase any remaining doses of the vaccine. It could also have made sure that every state had a plan for dealing with a vaccine shortage. And it could have helped to develop cell-culture techniques for flu-vaccine production as alternatives to the chicken eggs.

A mixture of ideological opposition and bureaucratic sclerosis has forestalled these actions. And, strikingly, three years of heightened concern about bioterrorism have done nothing to address the fundamental weakness of the US public-health system. Whether the threat is bioterror or flu, the lack of robustness in that system has never been more apparent. ■

Fishing for excuses

The message from researchers about the state of European fish stocks is consistent, but its delivery could be improved.

It is the same procedure every year: based on the advice of fisheries biologists across Europe, the International Council for the Exploration of the Sea (ICES) has again recommended a ban on cod fishing in the North Sea, the Irish Sea and west of Scotland.

As usual, the advice is likely to fuel a public outcry in the regions where fishermen live. And the European Union (EU), which sets annual catch quotas for fish, must walk a fine line between conserving the short-term future of the fisheries industry and the long-term survival of fish stocks. This annual event has already driven a wedge between scientists and fishing communities, a bad starting point for addressing the problem of fisheries conservation. Not only do fishing communities tend to dig in their heels, but conservationists sometimes exaggerate the need for draconian reductions in quotas.

Marine biologists often disagree over the best ways to conserve declining fish stocks worldwide (see *Nature* **419**, 662–665; 2002). And it isn't certain that fishing for threatened species — such as European cod — has to stop entirely, if stocks are to recover. This year, for example, cod stocks in the North Sea appear to have bounced back up by almost one-third, albeit from a worryingly low estimate of 35,000 tonnes, despite the fact that some fishing is permitted. ICES thinks

that no fishing should be allowed until stocks reach 150,000 tonnes.

But fish population dynamics is an uncertain business. Biologists' models don't work well if the stock in question is small, as is the case with European cod. Fish populations can fluctuate enormously from one generation to the next, depending on environmental conditions such as the weather, ocean turbulence and plankton availability. If the hydrographical conditions are favourable next season, cod stocks may recover further — but they could just as easily collapse. So from a precautionary point of view, it is understandable that scientists recommend fishing bans, even though they don't know for sure that continued fishing will wipe out European cod.

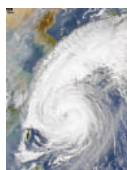
Such uncertainty does not release the EU from its obligation to optimize fisheries management. It can't do much about sea temperatures and ocean currents, but it can set sensible quotas on catches.

It probably makes sense that these quotas allow some cod fishing to continue, while reducing national fishing fleets and restricting the number of days that vessels are allowed at sea. Scientists can contribute to this process by acknowledging the uncertainties in their work. But it falls to fisheries managers and political leaders to persuade fishermen that quotas are fair and necessary. ■





Deep divisions?
Geneticists question
the validity of racial
groupings
p1026



Looking ahead
Japanese forecasters
seek better data
from typhoon
p1028



Hobbit unearthed
Dwarfed ancestor
discovered in
Indonesia
p1029



Time for bed
Tired jellyfish take to
the seabed to catch
40 winks
p1030

Science shares the limelight as election battle enters final phase

Erika Check and Geoff Brumfiel, Washington

This year's epic battle to secure the White House has bitterly divided America. But as the final showdown on 2 November nears, observers agree on one thing: science has for once played a prominent role in the contest.

Several key research issues found their way into speeches and news headlines during the campaign, including stem-cell research, nuclear-weapons proliferation, nuclear waste and the alleged politicization of scientific decisions under George W. Bush's administration.

But the topic that grabbed the nation's attention as the campaign reached its frenetic peak in October was the shortage of flu vaccines — an issue especially important to older Americans, the nation's most reliable voters.

The shortage made the news on 5 October, when California-based Chiron announced that some 48 million doses of flu vaccine made in Britain would be lost because of manufacturing problems. The shortfall was raised in the third and final presidential debate on 13 October, when Bush said he hoped that Canada could help replace the lost doses, but added "if you're younger, don't get a flu shot this year". Presidential challenger John Kerry made the most of the problem, alleging that Bush's healthcare plan for America was "don't get sick".

The candidates traded daily attacks on the issue for the best part of a week, with Bush claiming that it showed how lawsuits have scared vaccine makers out of the market. Kerry countered that the shortage revealed the damage done by Bush to the US healthcare system.

The United States has already seen two shortages of flu vaccine in the past four years, and numerous reports have warned that the country's vaccine manufacturing and distribution systems need a major overhaul. Whether that will now happen remains unclear. "The plus is that the issue will get publicized some more," says Frank Sloan, an economist at Duke University in Durham, North Carolina, who chaired a 2003 Institute of Medicine report on vaccine financing. But he adds that nothing he has heard from the candidates convinces him that either has a firm grip on how to fix the problem.



The third presidential debate saw George Bush (right) and John Kerry clash over flu vaccines.

Stem-cell research has also held a prominent position in the campaign, especially after the death of the actor and research advocate Christopher Reeve on 10 October. Kerry has pledged to overturn Bush's 2001 decision to limit federal funding for stem-cell research. His campaign staff believe that this has helped Kerry to make inroads with undecided voters who have relatives with diseases that might be treated as a result of stem-cell research.

Broader critique

Kerry has tried to broaden the stem-cell issue into a wider critique that Bush ignores scientific advice when setting national policy. "By blocking stem-cell research, President Bush has sacrificed science to ideology," Kerry told a crowd at Ohio State University in Columbus on 21 October — a message that has been amplified by a group of scientists travelling the country and delivering speeches detailing Bush's alleged abuses of science.

Physics-related issues have also made their mark. Nuclear non-proliferation took an unexpectedly prominent role in the first presidential debate on 30 September, when Kerry and Bush both identified it as their top priority in foreign policy. And the pair continued to clash over plans for a nuclear-waste dump — which Bush supports and Kerry opposes — at

Yucca Mountain in Nevada, one of the most keenly contested states in the election.

But the prominence of such issues won't necessarily help science, some observers say. They contend that scientists' backing for fewer restrictions on stem-cell research and for California's Proposition 71, which would spend \$3 billion on the field in that state, might send the message that the scientific community has a political agenda of its own.

"I think ultimately it's really bad for science," says Daniel Sarewitz, a science-policy analyst at Arizona State University in Tempe. "It introduces science as legitimately politicized by the scientific community." That could make it more difficult to argue that science should be allowed to operate without political interference, he says.

Henry Kelly, president of the Federation of American Scientists in Washington DC, and an adviser on scientific issues to Kerry's team, welcomes the involvement of scientists in this year's campaign. "If any citizen has something to contribute to the debate, they should be encouraged to step up," he says.

But John Marburger, Bush's science adviser, thinks that political action by researchers could erode bipartisan support for science. "I don't think there's any doubt that it's created a negative atmosphere for science," he says. ■

J.S. APPLEWHITE/AP

Ultrasound scans accused of disrupting brain development

Jim Giles, San Diego

The effect of ultrasound scans on brain development is to be investigated in a study on monkeys starting next month in the United States. The work has been prompted by unpublished research showing that ultrasound can disrupt the normal movement of cells through the brains of unborn mice.

The \$3-million study, which is funded by the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, will examine the effect of scans on the unborn offspring of around 50 rhesus macaque monkeys. Pasko Rakic and his colleagues at Yale University in Connecticut, will expose the monkeys to ultrasound at different times during pregnancy. Brain cells of interest will be tagged with a marker molecule before the scans, and the final position of the neurons will be assessed when the animals are killed after birth.

Rakic announced the study on 24 October at the annual meeting of the Washington-based Society for Neuroscience, held in San Diego. He will not reveal details of his preliminary mice study until the latest work has been published, but he says that the scans seem to interfere with the migration of neurons from the centre of the developing brain to the cortex — the outer layer of the brain that handles everything from movement to speech.

"The cells are slowed down and more spread out," he says. "Some of them are not getting to their final destination."

The movement of neurons in the developing fetus, on which Rakic did pioneering studies during the 1970s and 1980s, is known to be disrupted by certain viruses, genetic mutations and drugs taken during pregnancy. Studies have linked such disruption to a range of human conditions, including some forms of autism and learning difficulties.

Radiologists caution that more information is needed about the ultrasound dose used on the mice before any relevance to humans can be discussed. At high doses, for example, ultrasound causes a heating effect that damages tissue. "People have been studying the effect of ultrasound on development since the 1970s," says William O'Brien, a specialist in bioacoustics at the University of Illinois at Urbana-Champaign. "We've not seen anything when levels equivalent to those allowed for humans are used."

Geneticists struggle towards consensus on place for 'race'

Meredith Wadman, Washington

The thorny question of race is set to take centre stage at the annual meeting of the American Society of Human Genetics in Toronto, Canada, this week. Leading geneticists will present perspectives on how their discipline should deal with the fraught topic.

The geneticists' views will be published in a supplement to next month's *Nature Genetics*, and were to be the subject of a press conference that was due to take place in Toronto on 27 October. Their views took shape at what many participants characterize as a landmark meeting in 2003 at Howard University, a historically black university in Washington DC.

The scientists' overall message is that "the traditional categories of race and ethnicity are to some extent inhibiting researchers from identifying the real environmental and genetic causes of disease," says Myles Axton, editor of *Nature Genetics*. "The genetic differences that are of importance in disease are not necessarily distributed according to race and ethnicity." But the authors of the supplement, whose production was sponsored by the US Department of Energy, are sharply divided on how quickly they can afford to discard 'race' as a valuable category in their work.

Francis Collins, director of the US National Human Genome Research Institute, who contributed a commentary article, says that the connections between race, genetics and disease are ripe for serious scientific study.

Unlike some of his fellow authors, Collins does not think that the concept of race should be ditched. "Throwing it away might cause us to lose some of our best clues of the causes of disease, be they genetic or environmental," he says. However, he adds, race is "a lousy proxy" for specific risk factors and researchers should dispense with it as soon as they identify

pertinent factors. "It's a lot more powerful to test for sickle-cell carrier status than to look at the colour of somebody's skin," he says.

Human beings differ from each other at only about 0.1% of the genome's 3 billion bases, but it is not known whether or how those differences create disparities in disease prevalence, severity and drug response between different races. Some argue that studying this question will provide vital information for biomedical progress. Others say that this approach relies on unvalidated assumptions that race has a biological base, and plays into the hands of racists.

"Genetic variation research does not support the existence of human races. That is our overriding consensus," says contributor Charles Rotimi, acting director of Howard's human genome centre. "The only way we can get to the meaning of human genetic variation is by removing racial classification in biomedical research."

"We need to make changes," agrees Charmaine Royal, a geneticist at Howard and the prime mover behind the 2003 meeting. Study designs, she suggests, should reflect subjects' geographical location and ancestry, for example, rather than their race.

But Neil Risch, a geneticist at Stanford University, California, strongly disagrees. "It's extremely important to examine disease rates in different racial groups," he says. "There is no way to address inequities in health otherwise."

Joanna Mountain, a Stanford anthropological geneticist, says that nearly all researchers agree that, ultimately, race and ethnicity need to be discarded as categories in biomedical research. "But in the short term, it may be far more efficient to use them than to ignore them. Race and ethnicity are explanatory, even if it's unclear what they are surrogates for."



Stir it up: does race have a genetic basis that could aid biomedical research?

J. WIEDEL/ALAMY

Comet impact theory faces repeat analysis

Rex Dalton, San Diego

NASA is backing attempts to duplicate a published claim that a comet impact caused a mass extinction of species about 250 million years ago.

The space agency sent three scientists to China earlier this month to collect geological samples in an attempt to repeat the results of Luann Becker, a geochemist at the University of California, Santa Barbara, and her colleagues. Becker, whose research is funded by NASA, accompanied the group. The collected samples will be sent to ten laboratories for analysis early next year.

Deciphering the geochemical and seismic profiles of impacts millions of years ago that could have caused abrupt major changes in conditions on Earth is a highly contentious field. Inferences are drawn from geological samples taken around the globe, as researchers seek to identify impact craters that might be associated with a particular set of extinctions.

In 2001, Becker and her colleagues published an article reporting that the ratios of noble gases found in sediments in southern China were consistent with the theory that they originated in a comet or asteroid that hit Earth about 250 million years ago, at the boundary between the Permian and Triassic periods¹. It is known that up to 90% of Earth's species were wiped out at around that time. But no one has yet duplicated Becker's results, which have now been disputed in correspondence to *Science*.

In 2003, Becker's group published another article², which argued that pieces of meteorite found in Antarctica proved that



Luann Becker's claims that a comet caused mass extinctions 250 million years ago are disputed.

an asteroid impact had caused the Permian/Triassic extinction. Other researchers have disputed that result, contending that the meteorite fragments in question are not as weathered as they should be for that age.

And last week *Science* published two letters^{3,4} and a technical comment⁵ from Earth scientists criticizing a more recent article by Becker and her colleagues. That article located the asteroid's crater at a site off northwestern Australia⁶.

Jay Melosh, a geophysicist at the University of Arizona in Tucson and one of eight signatories to one of the letters in *Science*⁴, says that the Becker group "have deeply muddled the waters about what is going

on at the Permian/Triassic boundary".

But Becker says that she is looking forward to the new sample analyses for NASA, some of which she will conduct herself. She says her critics are being unreasonably aggressive. "This is science by intimidation," she says.

The critics say that they are driven by the lack of data backing up the original Becker papers. "They presented insufficient evidence of an impact crater or an age ascribed to it," says Paul Renne, a geochronologist at the University of California, Berkeley, and a signatory to the same letter to *Science*. "The latest *Science* paper undermines their credibility," says Renne, who argues that the data in the published paper do not support its conclusions. "A lot of researchers who were sceptical before are now sure Becker's group are wrong."

Michael New, a biophysicist who manages NASA's exobiology research programme, says that the agency learned earlier this year that some scientists were planning to do blinded studies that would repeat Becker's analysis of the Chinese samples, and decided to back it. The review is expected to cost about US\$100,000, with the results being published in mid-2005. "I thought it was a good idea to put together a consortium to figure out a consensus answer," he says. ■

1. Becker, L., Poreda, R. J., Hunt, A. G., Bunch, T. E. & Rampino, M. *Science* **291**, 1530–1533 (2001).

2. Basu, A. R., Petaev, M. I., Poreda, R. J., Jacobsen, S. B. & Becker, L. *Science* **302**, 1388–1392 (2003).

3. Wignall, P., Thomas, B., Willink, R. & Watling, J. *Science* **306**, 609 (2004).

4. Renne, P. R. *et al.* *Science* **306**, 610–611 (2004).

5. Glikson, A. *Science* **306**, 613 (2004).

6. Becker, L. *et al.* *Science* **304**, 1469–1476 (2004).

Spain's budget fails basic science, researchers charge

Monica Salamone, Madrid

Spanish researchers have lashed out at their new socialist government, accusing it of breaking pre-election pledges on science funding.

The government's first full budget since it came to power in March was released late last month — and researchers claim that it fails to deliver the major increase in science spending promised earlier this year.

In the run-up to the election, the socialists, led by José Luis Rodríguez Zapatero, had pledged to double Spain's €4-billion (US\$5.1-billion) annual research and development budget by 2008 (see *Nature* 428, 592; 2004). As part of this, researchers had hoped that spending would increase by 25% in 2005 as a first step towards that goal.

But the 2005 budget will see funding for scientists in universities and government

laboratories grow by just under 8%.

Prime Minister Zapatero's budget — which is subject to approval by parliament in December — would increase total research and development expenditure sharply.

But the bulk of this rise will go towards military work, including the development and testing of weapons systems. Scientists have called on the government to exclude the military component from its figures and to boost funding for civilian science and technology.

Much of the budget's increases consist of interest-free loans that will be made available mainly to businesses for their development work — something that won't help basic researchers.

"Companies or private foundations can ask for these loans, but how are we

scientists going to do it? It doesn't make sense," complains Joan Guinovart, president of the Confederation of Spanish Scientific Societies.

Guinovart and other scientists signed a statement in February that called for more state support for science, which they had hoped the new government would implement. And researchers this month issued another one saying that they are "enormously worried" by the government's budget plan. "We want a 25% increase in direct funds to research done in universities and public research centres," says Guinovart. ■

For more news and analysis go to
news@nature.com
www.nature.com/news

Deluge of typhoons may aid forecast models

David Cyranoski, Kobe

More than a hundred people have been killed by natural disasters in Japan in a little over a week. A magnitude 6.8 earthquake and severe aftershocks hit the western coastal region of the country, only days after an unusually violent typhoon ripped across the other coast.

Although seismologists continue to argue over the possibility of predicting earthquakes (see News Feature, page 1032), climatologists in Japan definitely see the 20 October typhoon as part of a trend. They are studying the footprint of the event to work out whether Japan is experiencing a short-term upward fluctuation in cyclone activity — or a more ominous, long-term increase linked to global warming.

Typhoons are like hurricanes but they emerge in the northwestern Pacific rather than the Atlantic. Both are intense tropical cyclones that form in low atmospheric pressure systems and give off winds travelling at more than 32 metres per second.

Last week's typhoon, which killed more than 80 people, was the 23rd such storm to be tracked by Japanese meteorologists in 2004. This is not an abnormally high number, but ten of these cyclones have come ashore, compared with an average of two or three a year since 1951, when modern records began. And four of them gave off winds of 40 metres per second when they hit Japan. The size of the typhoons and the amount of rain accompanying them is "super-abnormal", says Yoshio Kurihara, a typhoon specialist with the Frontier Research

Center for Global Change in Yokohama.

Kurihara says that higher average air pressures over the western North Pacific and higher sea surface temperatures are probably contributing to the severity of the typhoons.

But working out why that is will require greater knowledge of the structure of the typhoon, particularly of the "wall around the eye", where the wind is strongest and most of the storm's energy is released, says Kurihara. This wall varies in size, but its outer radius is normally about 40 kilometres and it can be 20 kilometres deep.



Heart of darkness: the cyclonic storm that hit Japan's Pacific coast last week took a heavy toll.

Unlike their counterparts in the United States, Japanese meteorologists lack aircraft that are capable of monitoring storm centres directly. They rely on a combination of satellite observation and ground-based monitoring to measure water vapour, wind speed, temperature and pressure around the centres. Kurihara and others are hoping that data from this year's storms can be compared with results from models of typhoons and the wider global climate system, and will lead to improvements in the models.

Akira Hasegawa, also of the Frontier centre, says that their model on the Earth Simulator — Japan's most powerful supercomputer — suggests that increased levels of carbon dioxide in the atmosphere will lead to lower or constant typhoon frequencies, with more rain. "If global warming continues, we will get typhoons with increasingly torrential rains," Hasegawa says.

This year's pattern would seem to confirm that prediction. But, Kurihara says, "it's too early to make any definite connections with global warming". Big typhoons were also observed in the 1950s, he notes.

This year's data will also be used to improve weather forecasts. Three days before the typhoon hit, the Japan Meteorological Agency made a landing-site prediction that was within 360 kilometres of the actual site and forecasted wind strengths that were only slightly weaker than those observed. "It was one of our better ones," says Nobutaka Mannoji, the head of the agency's typhoon division. ■

SEAWIFS, NASA/ORBIMAGE

Generic drugs allowed in global trial of AIDS therapy

Erika Check, Washington

Patients taking generic drugs will be permitted to participate in a worldwide trial that aims to find the best way to treat AIDS using antiretroviral medicines, US government officials say.

On 7 October, the scientific advisory board of Community Programs for Clinical Research on AIDS (CPCRA) voted to allow generic drugs to be used in its largest study. The trial, called Strategies for Management of Anti-Retroviral Therapy (SMART), is expected to include some 6,000 patients and to last a decade.

Officials at the US National Institutes of Health (NIH), which is funding the trial, say that the advice to include generic drugs has been accepted. "We're saying that the use of generic drugs in some countries in the study would not compromise the science," says Sandra Lehrman, director of the Therapeutics Research

Program at the NIH's Division of AIDS.

The United States does not allow its foreign aid to be used for buying generic drugs until they have been approved by the US Food and Drug Administration. Although the SMART trial does not fund the purchase of drugs, doctors in some countries said earlier this year that the NIH had decided that patients using generic drugs would not be able to take part in SMART. Doctors and AIDS activists criticized this decision, alleging that it was driven by the political clout of US drug companies, rather than by good medical practice (see *Nature* 431, 493; 2004).

The CPCRA says it will now consider enrolling patients who are taking generic drugs in the trial "on a country by country and drug by drug basis". Countries will be required to submit evidence that the drug they want to use has been approved by their own regulators. The CPCRA is now talking

to at least one country that would use generic drugs — Brazil — about its participation in SMART.

But it remains uncertain when, or if, other countries will join the trial, which will test alternative strategies for when patients should be treated with antiretrovirals. Enrolment in SMART is already behind schedule: only 40% of the patients expected to enrol this year have been signed up, and critics say the new policy leaves uncertainties in place that might increase delays.

But Martin Delaney of the non-profit AIDS foundation known as Project Inform, who is a member of the CPCRA advisory board, says the policy is a reasonable compromise. A broader approval for all generic drugs would be unwise, he says: "There are so many possible sources of drugs getting into the international AIDS drug market that it would be impossible to simply issue a blanket approval." ■



Evolution in focus: Peter Brown studies the skull of a new hominin species found in Indonesia.

Little lady of Flores forces rethink of human evolution

Rex Dalton

A new human-like species — a dwarfed relative who lived just 18,000 years ago in the company of pygmy elephants and giant lizards — has been discovered in Indonesia.

Skeletal remains show that the hominins, nicknamed 'hobbits' by some of their discoverers, were only one metre tall, had a brain one-third the size of that of modern humans, and lived on an isolated island long after *Homo sapiens* had migrated through the South Pacific region.

"My jaw dropped to my knees," says Peter Brown, one of the lead authors and a palaeo-anthropologist at the University of New England in Armidale, Australia.

The find has excited researchers with its implications (see News and Views, page 1043) — if unexpected branches of humanity are still being found today, and lived so recently, then who knows what else might be out there? The species' diminutive stature indicates that humans are subject to the same evolutionary forces that made other mammals shrink to dwarf size when in genetic isolation and under ecological pressure, such as on an island with limited resources.

The new species, reported in this issue of *Nature* (see pages 1055 and 1087), was found by Australian and Indonesian scientists in a rock shelter called Liang Bua on the island of Flores. The team unearthed a near-complete skeleton, thought to be a female, including the skull, jaw and most teeth, along with bones and teeth from at least seven other individuals. In the same site they also found bones from Komodo dragons and an extinct pygmy elephant called *Stegodon*.

The hominin bones were not fossilized, but in a condition the team described as being like "mashed potatoes", a result of their age and the damp conditions. "The skeleton

had the consistency of wet blotting paper, so a less experienced excavator might have trashed the find," says Richard Roberts of the University of Wollongong, Australia.

"Only the Indonesians were present at the actual moment of discovery — the Australian contingent had departed back to Oz," says Roberts. He credits Thomas Sutikna of the Indonesian Centre for Archaeology in Jakarta for the excellent handling of the samples. The success has inspired national pride at the centre, the researchers say. "This is very important for Indonesian society," says co-author R. P. Soejono.

The discovery is prompting increased scrutiny of sites on other Southeast Asian islands, both to look for more of the same species and to place it in context with *Homo sapiens* and *Homo erectus*, our closest relative. *Homo erectus* was found to have lived on the nearby island of Java as long as 1.6 million years ago; the team suggests that the Flores hominins may be their descendants.

Dating more bones could help determine whether the species was a short-lived branch of human evolution or survived for longer. Preliminary dating places it at about 70,000 years ago, but it may extend back 800,000 years. "We were hoping we might find a little hominin from that early," says author Michael Morwood, an archaeologist at the University of New England.

In the meantime, researchers are hoping to find DNA in the bones, which would help to clarify the relationships between species. DNA has previously been extracted from European Neanderthals living in the same time period. But they have so far failed to find DNA in the teeth of the *Stegodon* found in the same cave, says Brown. ■

Additional reporting by Michael Hopkin.

► www.nature.com/news/specials/flores

Novartis goes public with DNA data in bid to tackle diabetes

Jonathan Knight, San Francisco

The pharmaceutical company Novartis has entered a rare public-private partnership that will require it to place a mass of genetic data in the public domain.

The Broad Institute and Novartis Institutes for BioMedical Research, both in Cambridge, Massachusetts, announced that they have pooled resources to hunt for genes linked to adult-onset diabetes, the most common form of the disease. Novartis will contribute its diabetes research group to the project, as well as \$4.5 million over three years. The Broad Institute, headed by genome sequencer Eric Lander, will contribute its expertise in analysing genetic variation. The results are to be deposited on the Internet as they are generated, where they will be freely available. Novartis says it will share with the Broad Institute the rights to any potential drugs that emerge.

The research will investigate the genomes of several thousand Swedish diabetes patients and their families, from whom DNA samples have already been collected by Novartis and researchers at Lund University. The Broad team will search for small genetic variations that seem to correlate with the disease.

The large sample size should prove a boon to researchers. Diabetes is thought to result from a combination of genetic and environmental factors. Because researchers are looking for versions of genes that raise the risk of developing diabetes, rather than genes that definitively cause the disease, they need DNA sequences from a large number of people. "Without large patient samples, you can't distinguish signal from noise," says David Altshuler, who will direct the Broad Institute's share of the project.

It is rare for a private company to pay for a large public database. "To make it public is a good move for science, but it is unusual," says Sheldon Krinsky, a science-policy researcher at Tufts University in Medford, Massachusetts. Companies usually require a much more direct commercial benefit from such a partnership, he says.

The project is likely to generate far more data than the company can use, says Tom Hughes, head of diabetes research at Novartis Institutes. Releasing it will help the field as a whole, which in turn could lead to other commercial opportunities, he suggests. "We don't have to hold it all to ourselves to advance as a company." ■



Off the rails: a Japanese earthquake last weekend caused the first derailment of a bullet train.

Warning system for quakes still on track, say Japanese experts

Tokyo An earthquake of magnitude 6.8 tore up the western coastal region of Niigata in Japan on the weekend of 23 October. The tremor claimed at least 25 lives, and demonstrated its power by throwing one of Japan's bullet trains from its track — the first derailment since the trains came into service in 1964.

The train fortunately kept travelling in a straight path after derailing, as if it were “driving on an unpaved road”, says earthquake engineer Yutaka Nakamura of Tokyo-based System and Data Research. No one was hurt in the incident.

The derailment has thrown open a debate about the country's early-warning system for earthquakes, which can send out an alert faster than the waves of an earthquake in order to give some areas a vital few seconds of warning (see page 1032). Some critics wonder if such a system is worth the cost if it isn't fast enough to stop a train. But Nakamura, who helped to develop the system, points out that it is also critically useful to slow it down.

In this instance, the train slowed from 216 km h⁻¹ to 200 km h⁻¹ in its few seconds of warning. Although that may seem trivial, it is estimated to have cut 10% from the distance the train travelled after derailing — from 2.2 km to 2 km. This could in some instances prove life-saving, engineers believe.

Report calls for transgenic label on US maize exports

San Diego Transgenic maize imported into Mexico should be ground up and labelled as genetically modified (GM), according to a panel of scientists advising North American governments on trade issues.

This move would prevent the accidental flow of genes between GM crops and natural maize varieties, they say. The

United States and Canada currently export millions of tonnes of transgenic maize to Mexico.

The recommendation has been made to officials with the Commission for Environmental Cooperation, a body within the North American Free Trade Agreement. Scientists on the panel say they are frustrated by delays in publishing the report; it is now scheduled for release in mid-November, after the US presidential election. But the environmental group Greenpeace secured a confidential copy of the panel's recommendations and released them on 18 October, hoping to spur faster consideration of the advice.

A spokeswoman for the US Environmental Protection Agency said the panel's recommendations were flawed and in need of more rigorous peer review.

Russia gives green light to climate treaty

Munich The Kyoto Protocol on climate change can finally come into force, following a positive vote on 22 October by the Russian parliament. The protocol, drafted in 1997, will become a binding international agreement 90 days after Russia formally informs the United Nations of its ratification. This is expected to happen in the next few weeks.

Russia is the thirty-sixth industrialized country to agree to cut its emissions of six greenhouse gases, including carbon dioxide, by at least 5% below 1990 levels between 2008 and 2012. The United States and Australia have said that they do not intend to ratify the protocol.

Russia's decision ends a year-long debate (see *Nature* 431, 12–13; 2004). Environmental and scientific groups worldwide have hailed the move as a major breakthrough in international climate-change mitigation efforts, although most agree that more needs to be done to stem global warming.

Dozy jellyfish sink to a deep sleep

Sydney Jellyfish have been spotted ‘sleeping’ on the sea floor. Jamie Seymour's research team at James Cook University in Cairns, Australia, observed this behaviour by glueing radio-transmitters to box jellyfish (*Chironex fleckeri*) in the wild. From mid-afternoon until dawn, the team noticed, the jellyfish lay motionless (see picture). “They go completely catatonic,” says Seymour.

Seymour had seen box jellyfish napping in his lab before, but he assumed this was a quirk caused by the stress of captivity. No other jellyfish have been seen to sleep in tanks — though this may be a result of the unnatural environments found in aquaria.

Many sea animals, including fish, have

Schering stocks up with stem cells in Japan

Kobe In a vote of confidence for Japan's emerging biotechnology industry, Berlin-based company Schering has opened a research division in Kobe's biomedical park.

Major investment by foreign companies in Japan's biomedical industry is rare, but Günter Stock, a member of Schering's board of directors, says it is a wise move for the company. “The field of stem-cell biology could well first bear fruit in Asia,” he says. The company estimates that the centre will have an annual operating cost of about €10 million (\$12.8 million); it kept set-up costs down to €8 million by renting pre-existing research facilities.

The opening was celebrated with a two-day workshop on regenerative medicine, featuring presentations from several stars of the neighbouring Center for Developmental Biology. Stock says centres like these should provide fertile grounds for collaborations.

Home helpers chip in to the rise of the robots

Paris Robots haven't quite taken over, but their numbers are steadily increasing, according to a survey by the United Nations Economic Commission for Europe and the International Federation of Robotics.

Nearly a million industrial robots are now in service, according to the World Robotics 2004 survey. In the car-manufacturing industry, Japan and some European countries now employ one robot for every ten workers. Household robots are on the rise as well — about 600,000 autonomous lawn-mowers and vacuum-cleaners are in use, the report says.

The market for speciality robots — for use under water, in surveillance or demolition work, or for medical purposes, for example — is projected to increase several-fold over the next few years, thanks to falling prices.



periods of rest. The researchers say that sleeping makes sense for predators such as the box jellyfish, which may need a long break to recover after active bouts of hunting.

A seismic shift in thinking

Earthquake researchers in the United States have long shunned the word 'prediction'. But, thanks to improved data and a change in public perception, cracks are beginning to appear in their resolve. David Cyranoski tracks the debate.

At last month's meeting of the Southern California Earthquake Center in Palm Springs, a certain word was whispered in corridors or condemned with expletives in cocktail-party conversations. On slides during talks it was written only as the 'p-word'.

You wouldn't think the term 'prediction' could provoke such strong reactions. But for earthquake researchers, it's perhaps easy to see why it does. The early history of earthquake predictions featured scientists studying animal behaviour and watching the night skies for strange lights. Even when seismic studies came along, predictions were more often wrong than right. Disillusioned, and wary that false predictions would cause more damage than they would prevent, researchers — particularly in the United States — turned their backs on the word and the concept.

"There was a lot of bad science calling itself prediction," says seismologist Lucille Jones, who is in charge of the southern California area for the United States Geological Survey (USGS). "People wanted to dissociate themselves from it."

But prediction is coming back into researchers' vocabularies, if not into fashion. Most of the credit — or the blame, depending on your position — goes to Vladimir Keilis-Borok of the University of California, Los Angeles (UCLA), whose recent predictions ignited public concern and interest¹. UCLA's controversial press release describing his prediction of an earthquake in southern California attracted huge media attention. The quake never hit, but the episode resuscitated the p-word and brought the field into the media spotlight. "It's like we're doing experiments with the public looking over our shoulder," says Tom Jordan, director of the Southern California Earthquake Center (SCEC) at the University of Southern California in Los Angeles.

At the same time, researchers armed with a growing range of instruments and techniques are becoming more confident that their results are scientifically significant and useful. More than a billion dollars' worth of earthquake monitoring equipment in Japan, the United States and elsewhere is being

complemented by new statistical methods and theories. "The quality of the data has skyrocketed. People feel they are poised to make some real progress," says Jones.

In response to all this, the USGS is moving to re-establish the National Earthquake Prediction Evaluation Council, a committee charged with advising the director of the USGS on the merits of particular predictions. "We have a responsibility to be an honest broker in assessing predictions," David Applegate, senior scientific adviser at the USGS Earthquake Hazard Program, said at the Palm Springs meeting.

The council was first established in the late 1970s but has not appointed any new members in 12 years. The USGS has drafted a new charter that is slowly working its way through the Department of the Interior and other

bureaux. Applegate hopes the committee will be up and running by next spring. Also, a joint USGS-SCEC working group — called Regional Earthquake Likelihood Models — hopes to begin contrasting various forecast models for California by January 2005.

Hopeful harbingers

In the 1970s, enthusiastic support of earthquake prediction was less controversial. Following the discovery of plate tectonics, scientists had faith that the problem could be cracked, and in some places earthquake predictions were taken seriously. In China, the government evacuated Haicheng in February 1975, after scientists made a prediction based on changes in land elevations, groundwater levels, seismicity and animal behaviour. A magnitude 7.3 earthquake struck two days later, and the evacuation is credited with preventing 120,000 injuries and fatalities.

But failure followed this success. Just a year later, a magnitude 7.8 earthquake hit the city of Tangshan, killing 250,000 and injuring 164,000 people. There had been no prediction for that area.

Researchers came to believe that prediction was beyond their means, if not impossible. Rainfall, water levels, radon emissions, seismic waves, land deformations, geoelectric signals, cloud formations and catfish had all been studied as possible harbingers of

"After a bad quake, people want disaster mitigation. Now public attention is shifting back towards basic science." — Jim Mori



Deeply flawed: current knowledge doesn't allow us to predict the San Andreas fault's next shiver.

quakes, but a solid connection to the three golden variables — time, place and magnitude — remained elusive.

A double-whammy came with the California Northridge earthquake in January 1994 and the Japanese Kobe earthquake in January 1995. Neither fault region was seen as a threat, and the lack of concern showed in each area's poor building regulations. Both quakes were devastating. Researchers in the two countries most devoted to earthquake studies had missed their cues — assuming there were any to begin with.

Even more discouraging, an assembly of 1,224 Global Positioning System (GPS) stations and about 1,000 seismometers spread around the Japanese archipelago failed to spot any seismic hints of the magnitude 8.0 Tokachi-Oki earthquake that shook northern Japan last September. "There was no clear sign at all. It was a shock," says Ichiro Kawasaki of the Research Center for Earthquake Prediction at Kyoto University.

Experts today would call China's 1975 prediction, and others based on simple precursor events, good luck rather than good science. Of all the thousands of predictions ever made for quakes — most of them academic curiosities rather than attempts at disaster mitigation — some are bound to hit the nail on the head purely by chance. "It's like going to Vegas," says UCLA geophysicist David Jackson, who has won money from his colleagues by betting against predictions.



for example — may still prove impossible. But researchers now have a better understanding of the complexity of earthquakes, which may help to pinpoint places or times where emergency efforts should be focused. At California's San Andreas fault, for example, researchers are drilling down several kilometres to inspect a point on the fault line where quakes originate, to determine stress levels, temperatures, rock type and water content. This should provide a huge insight into earthquakes — in some cases at least.

Those intent on understanding how earthquakes happen are also excited by the recent discovery of two ways in which the deep Earth can release energy.

The strong silent type

The first of these has been dubbed the silent, or slow-slip earthquake³. Such disturbances originate 30 to 40 kilometres down, last between a day and a year, and can release the energy of a magnitude 7.0 earthquake, but more slowly and without ever being felt at the surface. Friction at these fault lines is greater than in the freely moving faults that allow tectonic plates to creep by each other smoothly, but less than that at patches where stress builds up and triggers a major quake. GPS is generally used to detect

these silent quakes at the surface. In Japan, the country with the biggest array of GPS devices, ten such events have been seen in the past decade, disproving critics' claims that they are a rare and insignificant anomaly.

The other oddity is a tremor whose seismic activity looks like that created by magma moving under volcanoes, but that occurs nowhere near a volcanic area. Beginning in September 2000, Kazushige Obara of the Japanese National

Research Institute for Earth Science and Disaster Prevention in Tsukuba saw this kind of seismic activity in three places in western Japan, far from any magma source that might create it⁴. The tremors were in active earthquake zones, known as subduction zones, where an oceanic plate slides under a continental one, but they were a new phenomenon. "It's the first new source of seismic waves discovered in 50 years," says Bill Ellsworth of the USGS in Menlo Park, California.

Obara suggests that these non-volcanic tremors are caused by water taken down with a subducting oceanic plate to a depth of some 30 kilometres, where it is so compressed that it forces its way into fractures deep in the



Shaken faith: no predictions were made for the devastating quakes that hit Northridge, California, (left) and Kobe, Japan (right).

In some countries, scientists battled on despite the bad news, but in the United States it became a liability to mention work on predictions. Researchers keen on the field say they had to look beyond the National Science Foundation and USGS for funding. They began speaking in terms of forecasts rather than predictions, using a term borrowed from meteorology that gives a wider margin for error.

Particularly after Northridge and Kobe, the public's attention shifted to reducing the damage from earthquakes, and away from attempts to anticipate them. "Science followed public interest," says Kyoto University

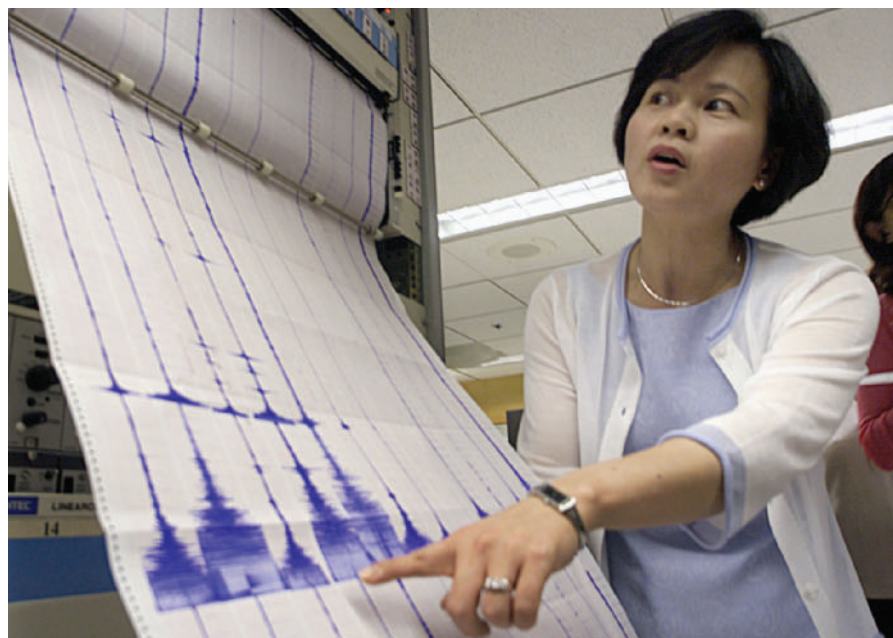
geophysicist Jim Mori, who worked at the USGS in the 1990s. Funding turned towards early-warning systems, for example, which spot the initial rumblings of a quake and send warnings across the city faster than the quake itself. Such systems, now established in Taiwan, Japan and Mexico City, can stop trains or shut down gas lines before disaster strikes. A similar network

is under consideration in California². "After a bad earthquake, people want disaster mitigation," says Mori. "Now public attention is shifting back towards basic science."

True prediction — of the sort that could be used to justify evacuating San Francisco,

"I didn't think my prediction method would work this well. I wish people would use it now."

— John Rundle



Sense and sensitivity: seismographs provide data that are vital for forecasting future events.

Earth's crust, or opens up new ones.

Both phenomena illustrate the complexity of earthquake generation, a welcome advance for researchers who knew that the simple models used for predictions were woefully incomplete. If the complex system could be understood, prediction might be possible, says Kawasaki, who tracked a silent quake⁵ in 1992. "These provide new perspectives that most people couldn't have imagined ten years ago," he says.

Silent quakes and non-volcanic tremors have even been found together in the Cascadia subduction zone, off the coast of the northwest United States and Canada⁶. Retrospective data analyses show that these have occurred in close to 14-month cycles for the past six years. "The Earth is beginning to look like it is behaving in an orderly way," says Ellsworth.

A recent quake at Parkfield in California also hints at a repeating system. This area was thought to have large earthquakes every 22 years. The latest quake, on 28 September, missed its predicted date by 15 years — but it did hit the right spot, reawakening debates about the cyclic nature of some quakes⁷.

Other researchers are looking for more complex patterns. John Rundle's group at the University of California, Davis, for example, is sifting through reports of small earthquakes in a search for hotspots likely to experience a major earthquake in the next ten years. His method assumes that seemingly chaotic patterns of magnitude 3 or 4 quakes can be used to reveal stress building up on a fault. When a threshold of stress is passed, a major quake is more likely, Rundle says.

Since Rundle published his results⁸ in February 2002, 11 earthquakes of magnitude 5 or greater have hit the California study area;

ten fell within range of his hotspots. "I didn't think it would work this well," he says. Rundle is also working on a map for Japan. On 23 October, a magnitude 6.8 quake hit Niigata — killing at least 25 people and injuring more than 2,000 — near one of Rundle's hotspots.

Rundle says his maps reduce the total area known to be seismically active to 24% of active fault areas, which would help to allocate resources for retrofitting bridges and other vulnerable infrastructure. "I wish people would use it now," he says.

Keilis-Borok also uses statistics and patterns to make predictions: his algorithms are derived from histories of large earthquakes. His most recent prediction concerned an earthquake of magnitude 6.4 or greater hitting a 32,000-km² area of southern California between 5 January and 5 September this year¹.

Shock tactics

Keilis-Borok's method⁹ has not been convincing, or even comprehensible, to many of his colleagues. Both his and Rundle's calculations require huge amounts of computation, leading some to charge that they are difficult for others to check.

Even if such long-term predictions were always correct, they would still leave public officials with the headache of deciding what to do with them — some fear that the panic caused by a quake alert might overshadow the benefits of an early warning.

The California Earthquake Prediction Evaluation Council, a local group that advises the state's governor on predictions, released a public notice on Keilis-Borok's prediction. It said his approach "had not been substantiated" and did not warrant any

specific action. But the same document called the approach "legitimate".

The resulting confusion showed the importance of providing the public with a clear message. The probability that there would be no earthquake, which Keilis-Borok put at 50%, never made it into the public perception, says Mark Benthien, the SCEC's director for communication, education and outreach. After 5 September passed, some people assumed the earthquake was just running late. Others thought the earthquake was set for 5 September exactly and ran out to get water the day before. The following day, one person wanted to know if it was all right to put picture frames back on the wall. "There is this idea that it's now over so we don't have to be prepared any more," says Benthien.

Ground rules

Governments are unlikely to embrace short-term predictions anytime soon, except perhaps in China, where 'official predictions' still occasionally hit the news. Even in Japan, where earthquake prediction studies abound and the word is not so feared, the government does not make official predictions, both to prevent panic, and out of a certain deference to the complexity of nature, says Kawasaki. There is an ethic that "research on prediction is a personal matter, but making predictions to the public must only be done with the consensus of the scientific community", he says. Clearly there is as yet no such consensus.

In the United States, the debate about the science, and the vocabulary used to describe it, goes on. At a press conference at the SCEC meeting, Jones encountered a frustrated journalist who demanded to know whether he should be using 'forecast' or 'prediction' in his stories. Why, he asked, was 'prediction' back on the menu after years of being told, and then convincing his editor, that 'forecasting' was more appropriate?

Many will say the debate is academic. Most earthquake researchers asked to define the difference between the words will sigh, and defer to a colleague. But Jackson pins it down: "Predictions are a subset in which probabilities become higher than normal for some reason — high enough to warrant some special action." If so, it is indeed a difficult word to use. But, with science, public perception and the media all pushing for a heightened awareness of the topic, the United States is getting ready to use it. ■

David Cyranoski is *Nature's* Asian-Pacific correspondent.

1. Cyranoski, D. *Nature* **431**, 117 (2004).
2. Allen, R. M. & Kanamori, H. *Science* **300**, 786–789 (2003).
3. Kawasaki, I. *Earth Planet. Space* **56**, 813–821 (2004).
4. Obara, K. *Science* **296**, 1679–1681 (2002).
5. Kawasaki, I. et al. *J. Phys. Earth* **43**, 105–116 (1995).
6. Rogers, G. & Dragert, H. *Science* **300**, 1942–1943 (2003).
7. Cyranoski, D. *Nature* **431**, 618 (2004).
8. Rundle, B., Tiampo, K. F., Klein, W. & Sá Martins, J. S. *Proc. Natl Acad. Sci. USA* **99**, 2514–2521 (2002).
9. Keilis-Borok, V. *Phys. Earth Planet. Int.* **145**, 75–85 (2004).

Bagging bursts, swiftly

The gamma rays spat out by dying stars last an instant. Tony Reichhardt reports on the fast-response satellite that hopes to capture them.

Ten years ago, science fiction could get away with speculating that gamma-ray (γ -ray) bursts were the exhaust trails of alien spaceships. And why not? No one had a sure explanation for these mysterious high-energy flashes, which briefly outshine everything in the γ -ray sky, and just as mysteriously flicker out, never to reappear.

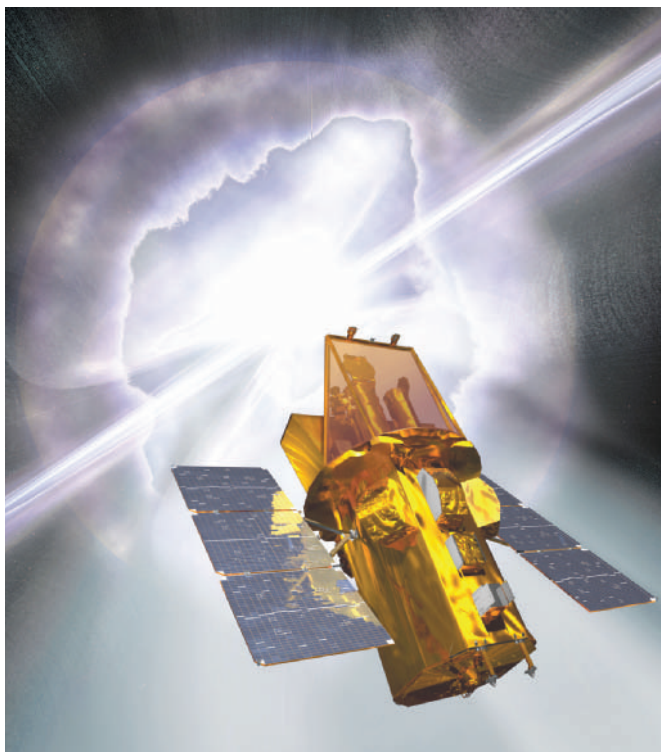
But the truth is equally dramatic. Most astronomers now believe these bursts are the dying gasps of massive imploding stars. Still, not all the mysteries have been solved. So a new NASA satellite called Swift is on the case, with launch scheduled for 8 November from Cape Canaveral in Florida.

Swift is not an acronym but a design philosophy. Speed is critical for studying γ -ray bursts, some of which last only a few milliseconds. The new satellite will be better than previous spacecraft at detecting the bursts, relaying their coordinates quickly to ground-based telescopes, then swivelling its own X-ray and optical/ultraviolet instruments to study the burst 'afterglow' that shines — albeit not as brightly — in other wavelengths.

This afterglow is key, because it tells observers how much energy is emitted at different stages of the explosion, and provides them with clues to its origins. There is a consensus regarding the source of longer-lasting γ -ray bursts, but the shorter bursts of less than two seconds remain mysterious, as no afterglow has been seen at any other wavelength. The team behind Swift hopes to fix that.

First noticed in the 1960s, γ -ray bursts are going off all the time, all over the sky. The Compton Gamma Ray Observatory catalogued more than 1,600 in the 1990s, but wasn't able to settle the key question of how far away they were. Then, in 1997, the Italian–Dutch spacecraft BeppoSAX saw an X-ray afterglow immediately following a burst. Optical afterglows were soon detected for other bursts, and these multi-wavelength observations confirmed that γ -ray bursts were very far away, in galaxies other than our own.

Theorists had trouble explaining how objects so distant could produce such



Glowing report: the Swift satellite will track gamma-ray bursts instantly.

astonishing amounts of energy — akin to our Sun's total lifetime output — in just a few seconds. But they reasoned that dying massive stars would be likely to expel jets of material at nearly the speed of light. The overall explosion didn't need to be so powerful if the energy was concentrated in jets shooting out from the star's poles. However, this means that we only see the small fraction of bursts whose jets happen to point towards Earth.

The long and the short of it

Further observations of γ -ray bursts and their afterglows led astrophysicists to lump these exotic objects into two categories. First and most common are the long bursts (more than two seconds) produced by large stellar explosions (hypernovae). The resulting 'collapsars' shoot out high-energy jets, although physicists disagree on whether the γ -rays come from the jets themselves or from a shell of material expanding rapidly away from the scene of the disaster.

Less common are the unexplained short bursts. These might represent a different type of collapsar, or they may be due to binary systems crashing into each other — a neutron star merging with another neutron star or a black hole. Scientists who model the physics of bursts would dearly love to have multispectral

data for the short bursts, and Swift should oblige them.

But classifying bursts as long or short doesn't capture their full diversity. Astronomers keep finding oddball γ -ray bursts with lower-than-typical brightness, or unusual energy distributions, that muddy the theoretical waters (see *Nature* **430**, 623; 2004). This leads veteran burst-hunters such as Gerald Fishman of NASA's Marshall Space Flight Center in Huntsville, Alabama, to argue that it's naive to try fitting all γ -ray bursts into two neat categories. "I definitely think we're in for some big surprises with Swift," he says.

Swift will have better speed, accuracy and sensitivity than today's best burst-finders: NASA's High Energy Transient Explorer (HETE-2) and Europe's International Gamma-Ray Astrophysics Laboratory (INTEGRAL). Swift's Burst Alert Telescope, with its large array of cadmium–zinc telluride detectors, will be the most sensitive

γ -ray imager ever flown. Mission scientists estimate they'll see between 100 and 150 bursts a year — although the number could climb higher because Swift will see fainter objects than previous telescopes have done. Astronomers hope it will smoke out a population of low-energy bursts that have escaped detection by less sensitive satellites.

They also expect the high-resolution images obtained by Swift's X-ray and optical/ultraviolet instruments to yield clues to burst origins. For nearby bursts, Swift should be able to pinpoint both the host galaxy and where in the galaxy the explosion occurred. If short bursts are located mainly in star-forming regions, that would favour the collapsar model, because hypernovae tend to be very young stars, unlike the much older neutron stars or black holes of binary systems.

Swift will also immediately relay precise coordinates for the burst to Earth, where dozens of observers — professional and amateur — will be poised to follow up. It should be possible to get telescopes viewing the scene within an hour, when the critical early afterglow will still be in progress. And this powerful combination of ground- and space-based observations, Fishman says, "will take us into uncharted waters".

Tony Reichhardt writes for *Nature* from Washington DC.

Scientists must conquer reluctance to speak out

When science is under political assault, keeping a dignified silence is counterproductive.

Sir — We read with some concern the views of M. J. Hsu and G. Agoramoorthy in Correspondence, that “Scientists and teachers should ignore politics” (*Nature* **431**, 627; 2004). They argue that scientists help society most effectively through teaching and research, rather than by taking part in election campaigns. In the current political climate in the United States, this well-intentioned argument represents a grave threat to both science and society.

The politicization of science threatens to undermine the value of science to society by obscuring scientific consensus and misleading policy-makers and the public. Although the threat is external — and most apparent in the suppression and manipulation of science by the Bush administration — the resolution is largely internal. More than 5,500 scientists have signed the Union of Concerned Scientists’ statement of protest, and more than 1,800 environmental scientists have signed a separate statement at www.scienceinpolicy.org. But it will take a greater outcry from the scientific community to bring this issue to the prominence it deserves. Scientists must step forward to protest against the manipulation of their results, or the obfuscation of accepted science will become an enduring tactic in political manoeuvring.

Already, scientific information is often clouded in the public arena. Evidence from competing expert witnesses in court cases, for example, makes it difficult for juries to decipher scientific evidence.

Attempts at journalistic balance similarly give equal weight to ideas that have unequal scientific support. This practice — which is neither good journalism nor an effective presentation of scientific knowledge — often creates the misconception that there is serious scientific debate about a particular issue when, in reality, there is virtually none.

For example, journalists gave roughly equal attention to the views of isolated scientists, including those funded by stakeholding industries, long after the wider scientific community reached consensus over the health threat posed by smoking and over the likelihood of human-induced climate change. In the former case, outcry from physicians and scientists finally penetrated the disinformation campaign by the tobacco industry (to society’s great benefit). Yet in the climate-change arena, the naysayers still have a significant voice despite the consensus against them.

Politicians increasingly employ a similar misrepresentation of science in public policy debates. If such manipulation is allowed to continue, scientists’ constructive

provision of unbiased, realistic assessments to policy-makers will be compromised.

Unfortunately, calling on scientists to defend their work from political manipulation bumps squarely against a deep reluctance among scientists to appear partisan. After all, the impartiality of science is largely responsible for the confidence most Americans have in scientific information. Scientists are legitimately concerned that advocacy may undermine the public perception that scientists are relatively apolitical and concerned primarily with facts. But what use is a voice that is held in high esteem but never raised?

We argue that the current assault on science sufficiently threatens the role of science in society to merit the risk of speaking out. Advocacy is less dangerous than sitting quietly on the sidelines while politicians and interest groups undermine the scientific method by perpetrating junk science.

Stephen Porder*, **Kai M. A. Chan†**,
Paul A. T. Higgins‡

*Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA

†Center for Conservation Biology, Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA

‡151 Hilgard Hall, University of California, Berkeley, California 94720-3110, USA

Why leave it to others to speak up about science?

Sir — Your correspondents M. J. Hsu and G. Agoramoorthy (*Nature* **431**, 627; 2004), seem to deny to scientists a right that lawyers, financiers, writers and even movie stars claim for themselves, which is direct involvement in political processes.

If this was ever a wise policy it is surely no longer, when science is so often the pawn of politics and individual politicians. The law on stem-cell research, for example, varies from one country to another according to political dogma. As a post-graduate researcher at a British university, I urge scientists to spare such time as they can afford to be involved with politics — as I have done myself, serving on a local council and even standing for parliament.

There is no good reason why lawyers and movie stars should have an exclusive right to debate science matters, any more than scientists should have the final say in the film industry or the law.

The fewer impenetrable membranes

with which scientists surround themselves and their work, the better.

Ian Flintoff

22 Chaldon Road,
London SW6 7NJ, UK

Knowledge is a good base to campaign from

Sir — M. J. Hsu and G. Agoramoorthy in Correspondence (*Nature* **431**, 627; 2004) imply that anybody choosing a career in science or education should disenfranchise themselves, and that a Nobel prize in a field other than peace equals a ban on political campaigning.

I wonder how the two correspondents think the celebrated American chemist Linus Pauling won his second Nobel: the 1962 peace prize? Surely, after being awarded his first one, in chemistry (1954), he should have stayed in his laboratory, setting his mind on loftier, ‘scientific’ matters, rather than meddling with such mundane business as international

relations and public affairs.

By campaigning to halt nuclear testing in the atmosphere, Pauling was both applying his technical expertise and following in the footsteps of others, including Albert Einstein (a pacifist who campaigned against nuclear weapons), who did not see their role in politics as limited by their Nobel prizes.

In the 1950s, Pauling paid the price for his peace campaigns, undergoing hours of interrogation and being refused the right to travel outside the United States. Although Einstein backed him, few other people dared to speak out during a period when an accusation of ‘un-American activities’ could cost them their livelihood or their freedom.

We do not have to repeat the lesson. Scientists should campaign in politics, and vigorously. I suspect Alfred Nobel, who instituted prizes not only in the sciences but also for peace, would approve.

Kaihsu Tai

Department of Biochemistry,
University of Oxford,
Oxford OX1 3QU, UK

Driven to diffraction

How Lawrence Bragg and his father used X-rays to solve crystal structures.

Light is a Messenger: the Life and Science of William Lawrence Bragg

by Graeme K. Hunter

Oxford University Press: 2004. 320 pp.

£35, \$59.50

Kenneth C. Holmes

As a child I was enthralled by William Henry Bragg's *Concerning the Nature of Things*, which answers simply the question recently put to me by my granddaughter: "What are atoms and molecules?". Both the author and his son William Lawrence Bragg were in their time resident professor at the Royal Institution of Great Britain in London, and both had an unusually well developed ability to communicate with school-children. Their joint Nobel prize, awarded in 1915, was for showing how X-ray diffraction could be used to determine the structure of crystalline substances. It is no coincidence that the title of William Bragg's book is a translation of *De rerum natura*, in which Lucretius set out his atomic theory of matter. However, Lucretius would have to wait nearly 2,000 years for the Braggs to show that he was right.

Lawrence was born in 1890 in Adelaide, Australia, where his father was a professor. He was a gifted pupil and became a very young member of the sixth form at St Peter's College. However, ignored by his older classmates, he was driven to finding solitary occupations, such as collecting and cataloguing sea-shells. At the age of 16 he proceeded to Adelaide University, where he took a degree in mathematics with first-class honours in 1908.

His father accepted an appointment as a professor at Leeds University, and in 1909 the family came to England. Lawrence entered Trinity College, Cambridge, taking first-class honours in the natural-science Tripos in 1912, and started his research under J. J. Thomson at the Cavendish Laboratory. His father had awakened his interest in Max von Laue's work on the diffraction of X-rays by crystals. Lawrence's studies of von Laue's diffraction patterns led him to postulate that zinc sulphide was based on a face-centred-cubic lattice, an amazing piece of insight. It was during this period that he formulated Bragg's law. Intuitively much simpler than the von Laue equations, it allows an estimate, by inspecting simple crystals, of how strong a particular X-ray reflection would be.

Lawrence started to work with his father in the summer of 1913. Although the older Bragg was still principally interested in X-ray spectra, his X-ray spectrometer also



Leading light: Lawrence Bragg, like his father, was resident professor at the Royal Institution.

provided a powerful tool for crystal analysis. After showing its power by analysing the structure of diamond, William continued to establish the relations between X-ray spectra and the K and L absorption edges, and Lawrence concentrated on interpreting crystal structures. It was the publication of their results in abridged form in 1915 that earned the two Braggs the Nobel prize for physics in 1915. At just 25 years of age, Lawrence was the youngest ever Nobel laureate.

During the First World War, Lawrence served as a technical adviser on sound ranging in France, where he made a number of friends, including R. W. James. Lawrence was appointed Langworthy professor of physics at Manchester University in 1919, and in 1921 he married Alice Hopkinson. He was neither a skilled lecturer nor a good administrator, however, and relied on James to keep the department running. But the remarkable science continued, with structures of the silicates and the optical theory of the diffraction of X-rays. The lab was abuzz with famous visitors. His father was then at the Royal Institution in London, presiding over Bill Astbury's unruly genius. Together with Kathleen Lonsdale and John Desmond Bernal, they were working out how

to do X-ray structure analysis of complex organic molecules. Between them, the Braggs had it sewn up.

After a year as director of the National Physical Laboratory in 1937–38, Lawrence became Cavendish professor of physics at Cambridge University (1938–53), finding Rutherford a hard act to follow, as he had at Manchester. Lawrence's avuncular style of lecturing was not to the liking of the students, and his crystallography did not please the nuclear physicists. Realizing that Cambridge did not have the resources to become an accelerator lab, he encouraged the study of radio astronomy and protein crystallography, which led to a plethora of Nobel prizes. His support for Max Perutz and his hopeless attempts to solve the Patterson function of haemoglobin was initially difficult to fathom, and later entailed tolerating Francis Crick's penetrating voice. Lawrence's memorable Edwardian epithet concerning Crick was that he was given to "doing someone else's crossword". All was forgiven when Crick and Jim Watson figured out the structure of DNA, however — not because Lawrence had any interest in biology, but because they beat his rival Linus Pauling. Nevertheless, Lawrence contributed a lot

to Perutz's subsequent success with protein structure. On a bizarre level, he was interested in crystal dislocations and, much to the amazement of his colleagues and first-year undergraduates, was able to simulate their motion with rafts of bubbles.

On retirement from the Cavendish, Lawrence became resident professor at the Royal Institution. There he built up a powerful group, led by David Phillips, that solved the first structure of an enzyme. In his lifetime, Lawrence saw X-ray crystallography grow from the seed he helped germinate to a method of solving the structures of the largest macromolecules.

The subtitle of Graeme Hunter's book refers to Lawrence's "life and science". The 'life' section is full of anecdotes and makes fascinating reading. Hunter captures the lonely schoolboy and tells of Alice Bragg — who some of us remember as a rather formidable justice of the peace — as a lively young flapper. He brings out Lawrence as an artist. Moreover, although Lawrence tried to avoid confrontation, his appointment to succeed Edward Andrade at the Royal Institution was accompanied by bad feelings and tension, which is fairly portrayed and analysed by Hunter.

The science is more of a problem. Most of it seems fairly accurate, but one or two sections reminded me of the description of the farm in Stella Gibbons' *Cold Comfort Farm*, in which the detailed geometrical descriptions resist synthesis. Hunter, who is not a crystallographer, must be commended for his brave attempt to put the science where it really belongs. However, his lack of a real understanding of diffraction theory shows up in numerous mini-howlers.

For example: " $n\lambda = 2d\sin\theta$... this was the famous Bragg equation. However, there was nothing really novel about it ... for a line grating, $2e\sin\theta = n\lambda$." Apart from the θ being different, Hunter misses the point first made by von Laue that diffraction from a three-dimensional lattice is subject to constraints not pertinent for a one-dimensional grating. The Bragg law imposes two conditions: specular reflection and the Bragg equation.

Hunter's lack of comprehension leads to an even bigger howler in Figure 0.2 in the introduction, which is supposed to help the lay reader. Hunter's putative Bragg reflections do not satisfy the Bragg equation, and moreover show that high-order reflections come out at low diffraction angles, and low-order reflections come out at high angles of diffraction; this is exactly the wrong way round. The strange thing is that a bit later, in Figure 2.7, he gets it right. The book, which could easily have been rescued by rigorous professional editing, is already in need of a second edition. ■

Kenneth C. Holmes is in the Department of Biophysics, Max-Planck-Institut für Medizinische Forschung, Heidelberg 69120, Germany.



Driving force: the New Madrid earthquakes moved the earth, but did they reshape geology?

Shaking up seismology

The Big One: The Earthquake That Rocked Early America and Helped Create a Science

by Jake Page & Charles Officer
Houghton Mifflin: 2004. 220 pp. \$24
Naomi Oreskes

In the winter of 1811–12, three major earthquakes struck an area of the North American mid-continent in rapid succession. According to eye-witnesses, the ground ruptured profoundly in numerous locations, lakes appeared where there had been none, and the mighty Mississippi River flowed backwards. The earthquakes, felt as far away as Montreal in Canada, affected an area of more than a million square miles. Their magnitudes have since been estimated at between 7.8 and 8.3, greater than the 7.6 of the famous San Francisco earthquake of 1906, making them among the most powerful quakes to strike the United States in recorded history.

The United States was then a young and sparsely settled country, and the theory of plate tectonics was far in the future, so there is no meaningful sense in which these earthquakes could have been considered "anomalous" at the time. Nonetheless, they are scientific anomalies now: the theory of plate tectonics explains large earthquakes as the release of stress built up as the Earth's crustal plates slowly grind past one another, but the quakes of New Madrid (to rhyme with Hagrid) occurred nowhere near a plate boundary.

If the theory of plate tectonics does not explain 'intra-plate' earthquakes, then what caused the New Madrid quakes? And why hasn't this conspicuous anomaly caused a crisis for the current theory? These are intriguing questions, and *The Big One* begins with the promise of answering them. Unfortunately,

that promise remains unfulfilled.

The book opens with a fast-paced description of the events of that winter and the background to European settlement in the region. It then changes tack: most of the rest of the book is a historical discussion of developments in the Earth sciences, leading to present-day theories of the origins of the New Madrid events. Sadly, this material is filled with factual errors and presents little that is not better treated elsewhere.

It would be tedious to recount the numerous mistakes and misrepresentations; a few will suffice to make the point. Lord Kelvin did not originate the idea that Earth was progressively cooling; that honour, if that's what it is, belongs to Georges-Louis Leclerc de Buffon, Immanuel Kant and Pierre Laplace. Isostasy — the theory that the Earth's crust sits in hydrostatic equilibrium on a denser substrate — is not the theory of glacial rebound; glacial rebound is merely one example of an isostatic effect. Alfred Wegener, the author of continental drift theory, did not die attempting to bring help to stranded members of his 1930–31 Greenland expedition, but on a trip to equip an inland observation station (see <http://www.awi-bremerhaven.de/AWI/geschichte/germanexpedition-e.html>). And no one in the 1920s pejoratively called Wegener's work "geopoetry" — that term was introduced later by the Dutch geophysicist J. H. F. Umbgrove as an approbative term for creative speculation, a concept later used to great effect by US geologist Harry Hess.

The authors' treatment of continental drift and plate tectonics is particularly beset by peculiar biases. They perpetuate the well worn but erroneous view that continental drift was rejected for lack of a causal mechanism, but in fact mantle convection was widely discussed in the 1920s and 1930s as a plausible mechanism. They credit the idea of mantle convection to seismologist Beno Gutenberg at Caltech in the 1950s, but its earliest prominent and credible advocate was the British geologist Arthur Holmes,

Science in Culture

From art to environment

Betty Beaumont's *Ocean Landmark* is in deep water.

Martin Kemp

Artists have long made large-scale interventions in the landscape. My house in Woodstock in Oxfordshire overlooks Blenheim Park, in which a magnificent landscape with a lake, a palladian bridge, rolling hills and clumped trees was sculpted in the 1760s by Lancelot 'Capability' Brown. Reshaping the land in this way requires patience: the scene progressively matures as the trees grow to full majesty and the ecology of the new topography organizes itself. More recently, 'land art' has involved the construction of huge artworks in specific locations, most notably the great *Spiral Jetty* constructed in 1970 by Robert Smithson at Rozel Point on the Great Salt Lake, Utah.

Betty Beaumont, an artist born in Toronto, Canada, but based in New York, follows this tradition of reshaping landscapes, but with key differences. Her interventions are directed specifically at social awareness, setting up environmental processes over long periods of time, rather than making monuments to be viewed in the time-honoured way. Indeed, the grandest of her long-term projects, *Ocean Landmark*, now almost 25 years old, cannot be readily viewed, as it lies deep in ocean water.

In 1978, Beaumont started work with a team of scientists to transform processed coal waste from a hydroelectric power plant in Ohio into inert rectangular blocks. Some 500 tons of the coal fly-ash blocks, 17,000 in number and cast at a concrete plant in Pennsylvania, were transported by barge in 1980 to a site on the continental shelf in the Atlantic Ocean, 40 miles from New York Harbor and 3 miles off Fire Island National Seashore. They were deposited on the sea floor to form a large mound.



Over the years, the austere blocks have been transformed into a lush reef, a rich ecosystem teeming with fish. Such has been the success of the project that it is listed as a 'fish haven' by the National Oceanographic and Atmospheric Administration.

Beaumont is fascinated by two time frames. The first consists of the ancient laying down of coal, the modern generation of power and the simultaneous production of waste. She completed the cycle with her team: "We took this material, transformed it and put it on the bottom of the sea at a planned depth so that life could develop, and it has sprouted an ecology that supports life, including plant life."

But is it successful as a work of art? If we are to define it as such, we have to stretch our definitions. She explains: "*Ocean Landmark* is an interdisciplinary work that at the time could only be described through other practices. It is known beyond an

'idea' as a real artwork. Although this artwork cannot be seen, each of us can visualize it."

Beaumont uses various strategies to meet our desire to see her work. One method, shown here, is to create a pile of scaled-down blocks as a surrogate for the underwater reef, but without, of course, the ecological accretions. Other strategies involve the sort of multimedia displays that natural-history museums use to portray aspects of nature.

Beaumont explains: "Current technology enables me to image this work in its life-giving, mature condition and in its entire form. Using global positioning satellite technology, the work can be located and images created through the use of underwater remote sensing and side-scan sonar. Coded in the images of the now-evolved underwater sculpture will be its progression as a sustaining environment for marine life and a thriving ecosystem."

Walking on the bank by Capability Brown's lake in Blenheim Park, I watch a pair of great crested grebe carrying fidgety chicks on their backs. Clearly Brown knew how to create bodies of water that were ecologically viable, although his main purpose was to delight the Duke and Duchess of Marlborough. In our era, Betty Beaumont takes this one step further, making viability the prime purpose of her art.

Beaumont's work is on show in the exhibition *Anima Mundi: Soul of the World* at the Herbst International Exhibition Hall, The Presidio of San Francisco, throughout October.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt. His new book, Leonardo, was recently published by Oxford University Press.

30 years before. More oddly, the authors approvingly discuss the scientific contributions of Maurice Ewing and Bruce Heezen of Columbia University, both of whom opposed plate tectonics, yet make no mention of Hess, a principal architect of the theory and the man most responsible for reopening the debate in the United States.

Small errors are most relevant when they add up to a big problem, and the big problem here is the underlying theme of the book. Encapsulated by its subtitle, the suggestion is that the New Madrid quakes helped to launch the specialism of seismology, perhaps

even the whole science of geology. The authors claim, for example, that geology in the early nineteenth century was "in its infancy"; that "most people who thought of themselves as scientists still believed generally in the history of the world as specified in the book of Genesis"; that many (if not most) geologists accepted the chronology of Archbishop Ussher that Earth was created on 23 October 4004 BC; and that Charles Lyell was the founder of modern geology.

These claims represent views that have long been discredited by professional historians. By 1811 there were well developed

empirical and theoretical frameworks for the Earth sciences, developed primarily in continental Europe but rapidly making their mark in Britain and the United States as well. Conversely, seismology had its early roots primarily in Italy, but developed as an organized scientific discipline in the late nineteenth century in Japan, India and Germany. In Japan, the constant threat of severe earthquakes in a densely populated country hemmed in by the sea provided strong motivation; in India, Richard Dixon Oldham's study of Indian quakes led to his discovery of P and S waves; and in Germany,

precision instrument-building led to the manufacture of good seismographs. New Madrid had little, if anything, to do with these developments. *The Big One* is a sloppy book, based on an erroneous premise. ■

Naomi Oreskes is in the Department of History, University of California, San Diego, La Jolla, California 92093-00104, USA.

Population biology on the wing

On the Wings of Checkerspots: A Model System for Population Biology

edited by Paul R. Ehrlich & Ilkka Hanski
Oxford University Press: 2004. 371 pp.
\$64.50, £40

Norman Myers

Extinction is the single irreversible feature that lies at the heart of the biotic crisis overtaking the planet. But it is not the extinction of species that counts most, even though we are in the opening phase of a species extinction spasm to surpass anything since the demise of the dinosaurs and associated species 65 million years ago. More significant even than the loss of species is the extinction of populations — the discrete aggregations of individual organisms that make up species. It is populations that form the basis for the diversity and abundance of species overall. A few species comprise just one population, but most have hundreds. Worldwide there are, crudely reckoned across all species, between 1.1 billion and 6.6 billion populations. We are consigning populations to eventual extinction at a rate many times higher than that for species. This is the hidden extinction crisis, overlooked by the public and our political leaders — and it receives much less attention than it might from many biologists.

Populations also provide the ecosystem goods and services that support human economies and societies. For instance, winged insects such as butterflies and bees serve as pollinators, and disrupting this role can cause long-term cascading effects throughout ecosystems. The mass extinction of populations is propelling us into a grossly destabilized environmental future.

Because biologists can study only a very small proportion of all species and an even smaller proportion of their populations, there is a premium on identifying a few long-term field studies of populations that can shed light on key questions of evolutionary biology. Such studies have examined Galapagos finches, Gombe chimpanzees, mountain gorillas and Serengeti lions. Regrettably, there had been no such study for invertebrates until the two editors of this book, Paul



Spotting a pattern: checkerspot butterflies can help us to understand other invertebrate populations.

Ehrlich and Ilkka Hanski, started studying populations of one of the better-known categories of butterfly, the checkerspots, in central California and southern Finland.

There are at least 20,000 known butterfly species in the world, but the checkerspots make up fewer than 400 of them, and a good many of these are endangered. They are among the best-studied populations of all invertebrates, and so are crucial for our understanding of the millions of invertebrates that make up the vast majority of all species. The two editors and 13 contributing researchers have sought to use their 40-plus years of intensive field and laboratory study “to create one population biological analogue to the well-known model systems in other biological disciplines, such as the fruitflies of classical genetics”. The result is a collaborative overview of model systems in population studies.

The book reviews a spectrum of the basic biology of checkerspots, including reproductive and larval biology, feeding patterns, population structure and dynamics, ecology and taxonomy. There are extended discussions of such issues as dispersal and migration, colonization, inbreeding depression, predation and parasitism, genetic differentiation, habitat fragmentation, threshold disturbances (especially by humans), climate and conservation biology. To cite the editors’ ultimate purpose, the major intellectual challenge of population biology “is understanding the functioning of natural populations — how they are distributed and structured, how and why their sizes change, and how they evolve”. In many respects, the book offers basic insights into the ecological and evolutionary dynamics of insect populations generally, not just of checkerspots, and thus forms a classic of modern biology.

The book provides lots of lessons for conservation biology. Many butterflies occupy

successional habitats, which are in transition from one ecological state to another. So studying their populations indicates how far they can adapt their lifestyles to human-disturbed landscapes. Like many butterfly species, checkerspots favour open country. Humans have been a potent force in converting forests into open landscapes, but regrettably many of these are pesticide-doused farmlands, overgrazed pastures, golf courses and treeless subdivisions — far from suitable habitats for butterflies.

Butterflies are a staple of summer gardens, parks and other landscapes. Yet about one-fifth of European butterfly species are threatened or vulnerable, and roughly one-seventh of those in the United States and Canada are at risk in certain areas or in the whole of their ranges. The path towards extinction can be rapid. The large blue butterfly (*Maculinea arion*) in Britain declined from some 30 populations with an estimated 100,000 individuals in the mid-1950s to just a single population of only 250 adults in the early 1970s, and to final extinction in 1979. Conversely, several UK butterfly species have expanded their ranges in recent years, ostensibly in response to global warming.

In the main, however, the prospect for many butterflies is not propitious. This book offers many clues on how we can improve that prospect. ■

Norman Myers is honorary visiting fellow at Green College, Oxford University, Upper Meadow, Old Road, Oxford OX3 8SZ, UK.

Correction

In Benno Müller-Hill's review of the book on Adolf Butenandt (*Nature* 431, 246; 2004), it was wrongly claimed that Otmar von Verschuer told colleagues in 1946 of his and Gunter Hillman's involvement in the analysis of blood samples from Auschwitz. In fact, von Verschuer disclosed this information in a written report to the DFG in 1944.

A wake-up call

How failing a PhD led to a strategy for a successful scientific career.

Bruce Alberts

One of my most important formative experiences as a scientist was very traumatic at the time. In the spring of 1965, I had finished writing my PhD thesis at Harvard University, in Cambridge, Massachusetts, and had purchased aeroplane tickets to take my wife Betty and our one-year-old daughter with me for a postdoctoral year in Geneva, Switzerland. Only one step remained — a meeting of my thesis committee to approve the granting of my PhD degree in biophysics. No one in recent memory had failed at this late stage. But to my great surprise, the committee failed me, specifying the need for more experiments that eventually required six more months of research.

This was, of course, a great embarrassment and a shock to my ego. There were the practical problems of having to remain at Harvard — our apartment had already been rented to the next tenant and my small family had nowhere to live. But most importantly, I was to spend the next few months struggling to answer two questions that would be critical for my future. What had gone wrong, and did I really have what it takes to be a scientist?

As an undergraduate working with Jacques Fresco in Paul Doty's laboratory at Harvard, I was handed a research project that proved to be very successful. My undergraduate thesis was quickly converted into two important papers in 1960. This largely undeserved success gave me a false image of how easy it would be to do science. It also enabled me to persuade Paul Doty to allow me to test my own theoretical model for the initiation of chromosome replication as the centre-piece of my PhD research.

According to my model, the sites at which DNA replication begins (now called replication origins) should be located at the two ends of each DNA helix in a chromosome. If this model was correct, the enzyme DNA polymerase should create a transient covalent linkage between the two complementary DNA strands at the tip of a chromosome (a 'DNA crosslink'). I began an extensive search in DNA genomes for crosslinks that were located near the sites where replication begins. None of the tests supported my particular model, but I did find other crosslinks in all of the chromosomes that I tested. I spent several years characterizing these mysterious and unexpected 'naturally occurring crosslinks', but even 40 years later, their structure and origin are still not understood (*J. Mol. Biol.* **32**, 405–421; 1968).

In retrospect, the shock of having my PhD thesis rejected in 1965 proved to be a



Bruce Alberts: 'failure' was a blessing in disguise.

critical step in shaping me as a scientist, because it forced me to recognize the central importance of the strategy that underlies any major scientific quest.

I had witnessed the frustration of scientists who were pursuing obvious experiments that were simultaneously being carried out in other laboratories. These scientists were constantly in a race. It had always seemed to me that, even if they were able to publish their results six months before a competing laboratory, they were unlikely to make truly unique contributions.

I had used a different strategy. My approach had been that of predicting how a particular biological process might work and then taking years to test whether my guess might be right. This was enormously risky. The good news was that I was carrying out experiments that were different from those being done by everyone else. The problem was that these tests could produce only a 'yes' or 'no' answer. If 'yes', I might be able to add something unique to the world's store of scientific knowledge. But if 'no', I would learn nothing of real value — in this case, I could eliminate just one of the many possible ways in which DNA replication might begin.

I wanted to continue to focus on how DNA is replicated for my postdoctoral work in Geneva. But what strategy should I

choose? The months of analysis triggered by the wake-up call of my PhD failure finally produced an answer. I would look for a unique experimental approach, but one that would have a high probability of increasing our knowledge of the natural world, regardless of the experimental results obtained.

After a great deal of soul-searching, I decided that I would begin by developing a new method — one that would allow me to isolate proteins required for DNA replication that had thus far escaped detection. I knew that the enzyme RNA polymerase, which reads out the genetic information in DNA, binds weakly to any DNA sequence — even though this protein's biologically relevant binding sites are specific DNA sequences. If the proteins that cause DNA to replicate have a similar weak affinity for any DNA molecule, I would be able to isolate them by passing crude cell extracts through a column matrix containing immobilized DNA molecules.

Arriving in Geneva in late 1965 with my PhD degree finally in hand, I found that by drying an aqueous solution of DNA onto plain cellulose powder, I could construct a durable and effective 'DNA cellulose' matrix. A large number of different proteins in a crude, DNA-depleted extract of the bacterium *Escherichia coli* bound to a column containing this matrix. Moreover, these DNA-binding proteins could be readily purified by elution with an aqueous salt solution. Using this new biochemical tool and a large library of mutant T4 bacteriophages obtained from Dick Epstein in Geneva, I discovered the T4 gene 32 protein after moving to Princeton a year later as an assistant professor. This proved to be the first example of a single-strand DNA-binding (SSB) protein, a structural protein that plays an important role in DNA processes in all organisms (see *Nature* **227**, 1313–1318; 1970).

The strategy of investing in method development and then using this new method for a major series of experiments would be employed over and over again during the next 25 years of my career as a research scientist. As a result, my laboratory almost never felt that it was in a race with other laboratories, and our successes were sufficient to satisfy both me and many of the graduate students and postdoctoral fellows who would join my laboratory. It seems strange to recall that we may owe all it all to one very unhappy PhD thesis committee at Harvard, nearly 40 years ago. ■

Bruce Alberts is the president of the National Academy of Sciences, 500 5th Street, NW, Washington DC 20001, USA.

Human evolution writ small

Marta Mirazón Lahr and Robert Foley

We are the only living species of the genus *Homo*. Given the startling results of a cave excavation in Southeast Asia, it seems that we coexisted with another species until much more recently than had been thought.

The fossils described elsewhere in this issue probably left no descendants, are not very old, and were found on a remote island. Despite this, they are among the most outstanding discoveries in palaeo-anthropology for half a century. The two papers concerned — by Brown *et al.*¹ and Morwood *et al.*² — appear on pages 1055 and 1087 of this issue, and respectively describe the fossils and their archaeological context. The find is startling. It is of a pygmy-sized, small-brained hominin, which lived as recently as 18,000 years ago, and which was found on the island of Flores together with stone tools, dwarf elephants and Komodo dragons. Discoveries don't get better than that.

The Flores fossils add a new and surprising twig to the hominin family tree, which diverged from the chimpanzee lineage about 7 million years ago. The first African hominins existed 7–1.2 million years ago, were 1–1.5 metres tall, walked upright on two legs (that is, were bipedal), and had chimpanzee-size brains. These early forms comprised as many as six genera and fourteen species, of which the australopithecines are the best known. By 2.5 million years ago, our own genus, *Homo*, had emerged, with its different body shape, slower growth, greater reliance on meat in the diet, and 'encephalization' — larger brains than expected for body size. These were the first hominins to make stone tools systematically and to colonize Eurasia. They include the familiar names of *H. habilis*, *H. erectus*, *H. neanderthalensis* and, finally, *H. sapiens*, which put in an appearance about 160,000 years ago. The new fossil is part of this *Homo* group (Fig. 1).

Flores lies to the east of Java, and was probably never connected to the mainland. The presence of 800,000-year-old simple stone tools first attracted attention in 1998 (ref. 3), raising the controversial possibility that *H. erectus* had produced them and had crossed major sea barriers to reach Flores. Now we have the announcement of the discovery of an 18,000-year-old hominin skeleton from a cave, Liang Bua, on Flores. Although this date is more than 140,000 years after modern humans evolved in Africa, more than 25,000 years after *H. sapiens* reached Australia, and about 10,000 years after the last known Neanderthal, the

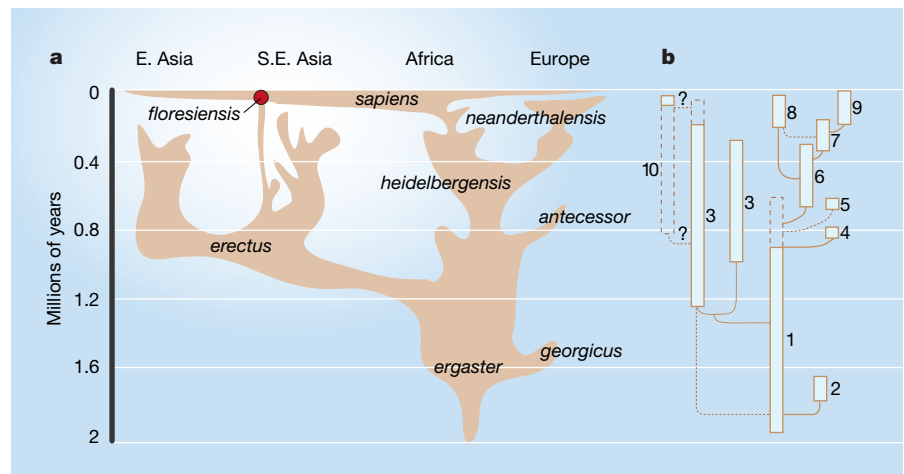


Figure 1 *Homo floresiensis* in the context of the evolution and dispersal of the genus *Homo*.

a, The new species as part of the Asian dispersals of the descendants of *H. ergaster* and *H. erectus*, with an outline of the descent of other *Homo* species provided for context. b, The evolutionary history of *Homo* is becoming increasingly complex as new species are discovered. *Homo floresiensis* (left) is believed¹ to be a long-term, isolated descendant of Javanese *H. erectus*, but it could be a recent divergence. 1, *H. ergaster*/African *erectus*; 2, *georgicus*; 3, Javanese and Chinese *erectus*; 4, *antecessor*; 5, *cepranensis*; 6, *heidelbergensis*; 7, *helmei*; 8, *neanderthalensis*; 9, *sapiens*; 10, *floresiensis*. Solid lines show probable evolutionary relationships; dashed lines, possible alternatives.

skeleton is that of a new species — *Homo floresiensis*. Its most remarkable features are its diminutive body (about a metre in height) and brain size (at 380 cm³, the smallest of any known hominin).

Homo floresiensis is a challenge — it is the most extreme hominin ever discovered. An archaic hominin at that date changes our understanding of late human evolutionary geography, biology and culture. Likewise, a pygmy and small-brained member of the genus *Homo* questions our understanding of morphological variability and allometry — the relation between the size of an organism and the size of any of its parts. Brown *et al.*¹ claim that the skeleton, designated LB1, represents a new species within the genus *Homo*. They believe that it may have been a female. They also conclude that it was a dwarfed descendant of Javanese *H. erectus*, and part of an endemic island fauna. But what other taxonomic assignments are possible?

Convergence — a process through which two species become more similar to each other than their ancestors were — is a strong evolutionary force⁴, and LB1, with its minute brain, could be a convergent Southeast Asian ape. But it evidently was an obligatory biped and had small canine teeth, key hominin

traits that, with the rest of its morphology, firmly place it within the hominin group⁵. Given its body and brain size, as well as some other features, could the remains be those of an australopithecine? Those features include bony reinforcements along the sides of the nose, thigh bones that were less obliquely aligned than ours (a trait essential for the way we walk and deal with gravity), and pelvic bones that were very wide, giving it a different overall body shape from ours. But the answer is again no. Most of LB1's other characteristics, such as the thickness and proportions of the skull, the flexion evident at the skull base, and the shape of the teeth, are derived traits of the genus *Homo*.

Could LB1 be a pygmy *H. sapiens*? Again, no. Compared with a human skull scaled to less than a third of full size, the LB1 skull differs in shape, robusticity and key features of the base. Furthermore, although human pygmies are short (1.4–1.5 m), they show very little reduction in brain size, probably because their small size is attained through mechanisms that curtail growth during puberty, when brains are already fully grown⁶.

In general terms, LB1's morphology groups it with *H. erectus*⁷. The name includes African and non-African hominins with

brains smaller than 1,250 cm³, which may be one species (*H. erectus*), or several (*antecessor*, *cepranensis*, *erectus*, *ergaster*, *georgicus*, *mauritanicus* and *soloensis*). Height among these 'erectines' is considered⁸ to range between 1.55 m and 1.78 m, and brain size between 650 cm³ and 1,260 cm³. The body and brain size of LB1 (about 1 m and 380 cm³) clearly indicate a major departure from the erectine extremes, while its peculiar combination of primitive and derived traits points towards the complex effects of dwarfism and its allometric consequences.

Island dwarfism is well known among mammals⁹. Released from predation pressure or constrained by restricted resources, and limited by population size, the phenomenon can be dramatic. Some examples can be truly extreme — for example, the one-metre-high fossil elephants, found on Sicily and Malta, which may have become dwarfed from a 4-metre ancestor in less than 5,000 years¹⁰. Indeed, remains of now-extinct primitive elephants (*Stegodon*), which had become dwarfed in relation to their mainland relatives, were found in the same deposits as LB1.

The dwarfism of *H. floresiensis* is also dramatic, resulting in the shortest adult *Homo*, and possibly hominin, known. Most significantly, the relative proportions of LB1's brain and body size (Fig. 2) indicate that the size reduction was more pronounced in the brain than the body, so a non-encephalized descendant evidently arose from an encephalized ancestor. This raises many questions about encephalization and hominin behaviour. Such questions aside, *H. floresiensis* is clear evidence that, in spite of their 'cultural niche', hominins were subject to the same evolutionary rules as other widespread mammals, with local isolation and small population sizes producing differentiation in size and form. This find strengthens the view that the genus *Homo* was probably much subdivided, resulting in a bushy human evolutionary tree. That view is itself consistent with the idea that the extreme climatic shifts of the past million years promoted population dispersal and isolation, and potentially resulted in instances of local evolution¹¹.

Necessarily, the discovery of *H. floresiensis* bears on the debate over the origins of modern humans — whether *H. sapiens* evolved in various regions throughout the world from *H. erectus* populations, or as a distinct and recent African species. Multiregional evolution requires the existence of large populations for long periods, with isolation being rare or absent so that the global species could evolve in a single direction. Palaeoanthropological and genetic studies have already done much to discredit this model, and *H. floresiensis* puts yet another (the last?) nail in the multiregional coffin. Not only did *H. floresiensis* evolve in the absence of gene exchange with other hominins, but

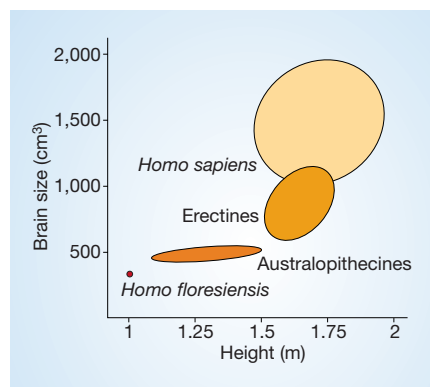


Figure 2 The relative brain and body size of *H. floresiensis*. The dimensions of the skull and skeleton (LB1) described by Brown *et al.*¹ fall well outside the extremes seen in *H. sapiens* and the 'erectines' (a range of hominin species, of which *H. erectus* is the most familiar). LB1 is closer in size to, but even smaller than, the australopithecines, of which the best known example is Lucy. On various anatomical grounds, however, Brown *et al.* believe that LB1 represents a dwarfed *H. erectus*.

no one can argue that LB1 contributed to our own species' genetic make-up.

Finally, accomplishing the sea-crossing that must have been necessary for the founding population to reach Flores adds to the baffling evidence for complex, supposedly 'sapient', behaviours among archaic hominins¹². And the behaviour of *H. floresiensis* itself, of course, remains elusive. Are the 800,000-year-old stones really artefacts? If so, does their date indicate when the taller ancestors of the dwarfed form arrived?

The archaeological evidence is controversial. The 800,000-year-old artefacts are simple, crudely flaked pebbles, similar to those found with Javanese *H. erectus*, as are some found at Liang Bua dating to more than 100,000 years ago. Only a few tools are associ-

ated with LB1. But thousands were found with the *Stegodon* skeleton in another sector of the cave: some are small flakes struck from radial cores; others consist of points, perforators, blades and possibly hafted microblades. Although Morwood *et al.*² attribute the production of all of these tools to *H. floresiensis*, elsewhere such implements are associated with *H. sapiens*, and their contrast with tools found anywhere with *H. erectus* is very striking. One could speculate that modern humans, who were dispersing across southern Asia between 100,000 and 50,000 years ago, may have made the tools, and come across these creatures. They may also have had a part in their ultimate extinction.

It is breathtaking to think that such a different species of hominin existed so recently. Brown *et al.*¹ point to the probability of similarly unexpected fossils being found in other isolated areas. For most of its 160,000-year history, *H. sapiens* seems to have shared the planet with other bipedal and cultural beings — our global dominance may be far more recent than we thought.

Marta Mirazón Lahr and Robert Foley are in the Leverhulme Centre for Human Evolutionary Studies, Department of Biological Anthropology, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

e-mails: m.mirazon-lahr@human-evol.cam.ac.uk
r.foley@human-evol.cam.ac.uk

1. Brown, P. *et al.* *Nature* **431**, 1055–1061 (2004).
2. Morwood, M. J. *et al.* *Nature* **431**, 1087–1091 (2004).
3. Morwood, M. J., O'Sullivan, P. B., Aziz, F. & Raza, A. *Nature* **392**, 173–176 (1998).
4. Conway Morris, S. *Life's Solutions: Inevitable Humans in a Lonely Universe* (Cambridge Univ. Press, 2003).
5. Wood, B. & Richmond, B. G. *J. Anat.* **197**, 19–60 (2000).
6. Shea, B. T. & Bailey, R. C. *Am. J. Phys. Anthropol.* **100**, 311–340 (1996).
7. Anton, S. Yb. *Phys. Anthropol.* **46**, 126–170 (2003).
8. McHenry, H. M. *Am. J. Phys. Anthropol.* **87**, 407–431 (1992).
9. Foster, J. *Nature* **202**, 234–235 (1964).
10. Lister, A. *Symp. Zool. Soc. Lond.* **69**, 277–292 (1996).
11. Lahr, M. M. & Foley, R. A. *Yb. Phys. Anthropol.* **41**, 137–176 (1998).
12. Foley, R. A. & Lahr, M. M. *Cambr. Archaeol. J.* **7**, 3–36 (1997).

Astronomy

Tycho's mystery companion

David Branch

A famous sixteenth-century supernova, seen by Tycho Brahe, is still a hot topic. The stellar explosion might have been initiated by a companion star — and modern astronomers have at last identified it.

On 11 November 1572, Danish astronomer Tycho Brahe looked up at the constellation of Cassiopeia and saw a bright new star. In fact, what he saw was the death of a star — a supernova. The appearance of this 'star', now known as Tycho's supernova, refuted the Aristotelian immutability of the heavens, and might have been the inspiration¹ for the celestial portent in the opening scene of Shakespeare's *Hamlet*. The remnant of the supernova,

10,000 light years from Earth, now glows at radio and X-ray wavelengths, owing to the strong interaction between the high-velocity matter ejected from it and the interstellar gas that was swept up in the cataclysm. The event is thought to have been a type Ia supernova — the complete disruption of a white-dwarf star provoked by the transfer of mass from a close binary companion. More than four centuries later, Ruiz-Lapuente *et al.*² (page 1069) claim to have identified the

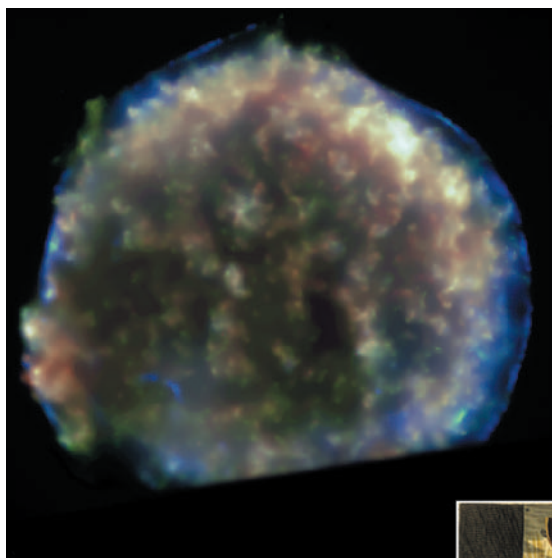


Figure 1 Tycho's supernova. In 1572, Tycho Brahe (below) recorded the appearance of the supernova that now bears his name. This image from the Chandra X-ray Observatory shows the expanding remnant of the explosion. The colours represent different X-ray energies; the high-energy shock wave appears as a blue outline. Tycho's supernova is thought to belong to the 'type Ia' class: the death of a white dwarf following the absorption of mass from a companion star. Ruiz-Lapuente *et al.*² may now have found that companion star.



long-sought companion star of Tycho's supernova (Fig. 1).

The observed properties of type Ia supernovae are highly homogeneous, one such property being the maximum luminosity, which exceeds that of a billion Suns. As a result, these supernovae make exquisite 'standard candles' for cosmology, because they can be seen from afar and their apparent peak brightness can be used to reliably infer their distance. In recent years, they have been observed at ever greater distances from Earth, and, because the speed of light is finite, this means that we are looking ever deeper into the past. The best-observed type Ia supernovae are in relatively nearby galaxies, at distances of tens of millions of light years; observations of more remote ones have been used to determine the value of the Hubble constant³, a measure of the recent expansion rate of the Universe. The apparent brightnesses of very remote type Ia supernovae, at distances of billions of light years—including some that erupted even before the Sun and Earth formed—have revealed that the cosmic expansion is not simply decelerating through the effects of gravity, but that it is accelerating, driven by some kind of 'dark energy'^{4,5}. The profound cosmological implications of this are the motivation for astronomers to strive to better understand this class of supernova.

The progenitor of a type Ia supernova is a white-dwarf star, which is composed mainly of carbon and oxygen (like the white dwarf that the Sun will become). The white dwarf accretes matter from a close companion star and contracts, thereby increasing its temperature and density, until it approaches the maximum stable mass for a white dwarf (which is 1.4 times the mass of the Sun). At this point a violent thermonuclear instability releases enough fusion energy to blast the white dwarf apart, at a few per cent of the speed of light. Nuclear reactions also burn almost half of the white-dwarf mass to a single isotope, radioactive nickel-56, the

decay of which, through cobalt-56 to stable iron-56, provides a delayed input of energy that keeps the expanding ejecta hot and explains why the supernova shines so brightly. Astronomers have great confidence in this much of the story⁶.

But they are less confident about the nature of the binary system that existed before the supernova. The white dwarf's companion may have been a normal star: this could have been a 'main-sequence' star (those powered by the fusion of hydrogen into helium), a more evolved 'subgiant' (whose hydrogen fuel is running low), or a red giant (whose hydrogen fuel is exhausted in its core). Although shaken up by the explosion, this star would have survived, running away from the site at its pre-supernova orbital velocity. But another possibility is that a type Ia supernova is produced when a pair of white dwarfs coalesces to form one rapidly rotating, overmassive white dwarf. This will either explode as a supernova or (perhaps more likely⁷) collapse to form



100 YEARS AGO

Is his letter of last week detailing his most interesting experiments on cross-bred maize, Mr. R. H. Lock makes the following statement:—"I see from the published account of a recent discussion at the Cambridge meeting of the British Association that the facts of Mendelian segregation are still disputed by the biometric school of evolutionists." Now it is easy to make a general statement about some vaguely defined group of men, and I have no right to speak for biometricians as a body. But as inventor of the term *biometry*, I may perhaps be allowed to say what I understand by it as a science, and to restate what I said with some emphasis at the Cambridge meeting. Biometry is only the application of exact statistical methods to the problems of biology. It is no more pledged to one hypothesis of heredity than to another, but it must be hostile to all treatment which uses statistics without observing the laws of statistical science. The criticism which has been published in *Biometrika* upon Mendelian work has attacked its too frequent want of method and of logic, and I think no one can have read recent literature without seeing that the criticism has been effective in its aim.

Karl Pearson

From *Nature* 27 October 1904.

50 YEARS AGO

Lord Boyd Orr, summing up the symposium ["The Numbers of Man and Animals"], said that he welcomed all the developments there might be in sanitation and preventive medicine, on one hand, and contraception on the other. To a large extent they cancel one another out, but he expected the world to hold perhaps 4G ($G = 10^9$) people by 1980 and 5 or 6G by the end of the century... The difficulties that he foresaw were political rather than scientific. Physics appears to governments to be much more useful than biology both in war and for making money; it therefore gets the lion's share of research endowment. People had got on very well in the past without jet planes and hydrogen bombs, but they could not get on without food, and he looked forward to an era of agrarian abundance in which we no longer galloped through irreplaceable resources with our present abandon but farmed wisely and depended for our energy supplies primarily on the inexhaustible flow from the sun.

From *Nature* 30 October 1954.

a neutron star. If the configuration does explode, there is no surviving companion.

The search for the erstwhile companion star to Tycho's supernova has been a challenging task. Ruiz-Lapuente *et al.*² have probed deep-space images from ground-based telescopes, looking for stars that are close to the centre of the present extended remnant of the supernova and at about the right distance from Earth. From data collected at optical wavelengths, they have calculated the velocities of candidate stars along the line of sight (radial velocities); and, from Hubble Space Telescope images, the velocities perpendicular to the line of sight (tangential velocities). Velocity is the key because, owing to the differential rotation of our Galaxy, stars in the direction of Tycho's supernova and at a similar distance have a characteristic velocity relative to the Sun. The data obtained so far reveal only one star that is at the right distance and that has an unusual velocity: a star, not much unlike the Sun, that is moving at about the orbital velocity expected of a mass-donating solar-type companion.

The discovery of Tycho's companion will intrigue many for the sheer historical

significance of the 1572 supernova. It will excite supernova enthusiasts because of the new information it provides about the nature of progenitor binary systems for type Ia supernovae. The data obtained by Ruiz-Lapuente *et al.* exclude the possibility that the companion was a red giant. If we accept that the companion has been identified, we now know for the first time that not all type Ia supernovae are produced by the coalescence of white dwarfs. Further observations of the putative companion — to search, for example, for signs that the companion's atmosphere has been contaminated with ejecta from the supernova — are eagerly awaited. This would clinch the case that Tycho's companion has been found at last. ■

David Branch is in the Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA.

e-mail: branch@nhn.ou.edu

1. Olson, D. W., Olson, M. S. & Doeschner, R. L. *Sky & Telescope* 68–73 (November 1998).
2. Ruiz-Lapuente, P. *et al.* *Nature* 431, 1069–1072 (2002).
3. Saha, A. *et al.* *Astrophys. J.* 562, 314–335 (2001).
4. Riess, A. G. *et al.* *Astron. J.* 116, 1009–1038 (1998).
5. Perlmutter, S. *et al.* *Astrophys. J.* 517, 565–586 (1999).
6. Hillebrandt, W. & Niemeyer, J. C. *Annu. Rev. Astron. Astrophys.* 38, 191–230 (2000).
7. Saio, H. & Nomoto, K. *Astrophys. J.* 500, 388–397 (1998).

other three are the most relevant to assess progress towards the 2010 target, being direct measures of the state of biodiversity: habitat indices, measuring trends in extent of selected biomes, ecosystems and habitats; population indices, measuring trends in abundance and distribution of selected species; and the marine trophic index, measuring changes in the mean food-web level of fisheries landings.

Each of these indices has advantages, but also limitations. Habitat and population indices have been proposed as particularly valuable in that they respond quickly to changing biodiversity⁵. Habitat indices are also geographically representative, but sample a coarse ecological resolution, which does not reflect finer levels of biodiversity organization (such as species) very well⁶. Global population indices, although they have a fine ecological resolution, are highly biased at present towards better-known (frequently biodiversity-poor) regions. The Living Planet Index⁷, for example, incorporates few tropical data sets, and those data that it does use may have been collected for populations already known to be declining. Like population indices, the trophic index⁸ has a fine ecological resolution but is not comprehensive: it is based on data for aquatic ecosystems only, and is specifically a measure of the impact of a particular threat (fisheries).

The Red List Index proposed by Butchart *et al.*² fills a portion of 'biodiversity indicator space' that complements the indicators already being tested. The index measures changes in overall extinction risk for all species, worldwide, in an entire class of organisms. For this first application it is tested for all birds⁹. It therefore has both fine ecological resolution and comprehensive geographical representativeness. The cost is a lack of the sharp temporal resolution achieved by habitat, population and trophic indices. This said, the Red List Index is still sensitive enough to reveal a serious increase in extinction risk for birds between 1988 and 2004. Further, it is robust enough for subdivision by biogeographical realm, habitat and taxonomic group. This shows, for example, particularly devastating increases in aggregate extinction risk for Asian species (Fig. 1) (due to deforestation, in Indonesia in particular), and for albatrosses and other pelagic seabirds (due to the increase in long-line fisheries).

The paper introduces several clever innovations. The data underlying the Red List Index are derived through quantitative, transparent and repeatable assessment of extinction risk under the IUCN Red List criteria¹⁰. Further, Butchart *et al.* carefully classify reasons for changes in Red List assessments, and introduce a method of 'back-casting' the index. This removes a potential artefact resulting from the many

Conservation biology

Biodiversity barometers

Thomas Brooks and Elizabeth Kennedy

The Red List Index is a new indicator of species' extinction risk. It will make a major contribution in measuring the success of an internationally agreed aim to slow biodiversity loss by 2010.

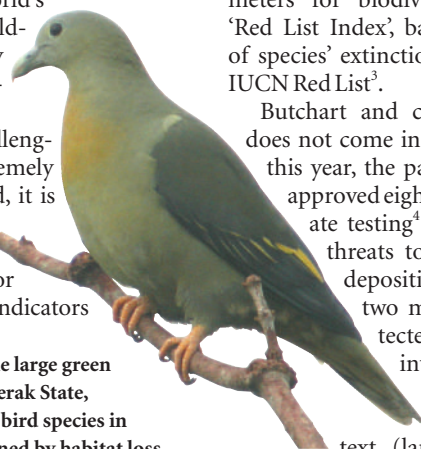
In April 2002, the 188 parties to the Convention on Biological Diversity agreed to reduce the rate of biodiversity loss significantly by 2010. This commitment covers all of the world's nations except Andorra, Brunei, Iraq, Somalia, Timor-Leste, the United States and the Vatican, and was endorsed by the World Summit on Sustainable Development later that year¹. The 2010 target is a ground-breaking advance, in that the world's governments are holding themselves publicly accountable for biodiversity conservation.

The target is also challenging. First, it will be extremely hard to achieve. Second, it is not even obvious how, in 2010, we will be able to determine whether or not it has been reached. Indicators

for doing so will effectively require the use of existing data sets, because assessment of a change in the rate of biodiversity loss will require a minimum of three temporal data points. They will also need to have comprehensive geographical extent but fine ecological resolution. Further, they should respond rapidly to changes. Writing in *PloS Biology*, Butchart *et al.*² make a major contribution towards the development of such barometers for biodiversity in proposing a 'Red List Index', based on the assessment of species' extinction risk provided by the IUCN Red List³.

Butchart and colleagues' contribution does not come in a vacuum. In February this year, the parties to the convention approved eight indicators for immediate testing⁴. Two of these measure threats to biodiversity (nitrogen deposition and water quality), two measure responses (protected area coverage and international conservation funding), and one measures cultural context (language diversity). The

Figure 1 In decline: a female large green pigeon (*Treron capellei*), Perak State, Malaysia. Like many other bird species in Asia, this pigeon is threatened by habitat loss.



changes to the Red List due to changing knowledge rather than to changing levels of extinction risk¹¹. They also use these back-cast data to calculate error around the current Red List Index value due to genuine but as yet undetected changes in status. Finally, they provide a weighting for the index based on extinction probabilities, through which it becomes largely driven by actual extinctions. Although this gives narrower biodiversity coverage (because the contribution of most species towards this weighted index is negligible), it may provide a better measure of the loss of irretrievable genetic diversity.

As an immediate step for the assessment of progress towards the 2010 target, we suggest that the next meeting (in February 2005) of the Subsidiary Body for Scientific, Technical and Technological Advice to the convention should add the Red List Index to the other eight indicators already recommended for immediate testing. The index then needs expansion in coverage across other taxonomic groups. All mammal and amphibian species are currently being assessed, and Red List Indices for these will be available by 2010. Efforts to assess reptiles, fish and plants are also in progress but will require significant support to produce timely results. Moreover, a sampled Red List Index is under development, with the aim of providing an index that is taxonomically as well as geographically comprehensive. Similarly, the other eight indicators proposed for testing need considerable work to better inform progress towards the 2010 target.

Although we need to be able to measure progress (or lack thereof) towards the 2010 target, the more pressing need is to ensure funding and implementation of those conservation activities needed to actually achieve a significant reduction in the rate of biodiversity loss. Ultimately, an improvement in the Red List Index will require urgent investment in the conservation of species facing a high risk of extinction and of the habitats where they occur.

Thomas Brooks and Elizabeth Kennedy are at Conservation International, 1919 M Street NW, Washington DC 20036, USA.

e-mail: t.brooks@conservation.org

- Kaiser, J. *Science* **297**, 1785 (2002).
- Butchart, S. H. M. *et al.* *PLoS Biol.* **2**(12), e383 (2004).
- www.redlist.org
- www.biodiv.org/2010-target/indicators.aspx
- Balmford, A., Green, R. E. & Jenkins, M. *Trends Ecol. Evol.* **18**, 326–330 (2003).
- Araújo, M. B. *et al.* *Ecography* **24**, 103–110 (2001).
- Loh, J. *Living Planet Report 2002* (WWF Int., Gland, Switzerland, 2002).
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Jr *Science* **279**, 860–863 (1998).
- BirdLife International *Threatened Birds of the World 2004* (BirdLife Int., Cambridge, UK, 2004).
- IUCN *IUCN Red List Categories and Criteria: Version 3.1* (IUCN — World Conservation Union, Gland, Switzerland, 2001).
- Cuarón, A. D. *Nature* **366**, 118 (1993).

Solar physics

Spots from rings

Paula J. Reimer

An ingeniously constructed record of sunspot activity shows that the current episode is the most intense for several thousand years. But that does not let us off the anthropogenic hook of global warming.

Dark spots or 'blemishes' on the face of the Sun (Fig. 1) were recognized from the early seventeenth century, and have since been identified as places where strong magnetic fields emerge from the Sun's surface. Data on sunspot numbers provide the longest observational record of solar activity. But that record is too short to document changes in activity occurring on timescales longer than the recognized cycles of 11- and 88-year periods, or to support claims for a connection between solar activity and Earth's climate on centennial to millennial timescales.

On page 1084 of this issue, however, Solanki *et al.*¹ describe how they have produced a reconstruction of sunspot number for the past 11,000 years. They have done so by connecting a series of models based on well-established physics, taking as their data the concentration of the carbon isotope ¹⁴C found in tree rings, which provides windows on atmospheric and solar trends at known points in time.

The reconstruction shows that the current episode of high sunspot number, which has lasted for the past 70 years, has been the most intense and has had the longest duration of any in the past 8,000 years. Based on the length of previous episodes of high activity, the probability that the current event will continue until the end of the twenty-first century is quite low (1%).

Each model used in the reconstruction makes a step in connecting the tree-ring ¹⁴C record² to sunspot number using parameters that were fixed by independent measurements (direct or indirect). Carbon-14, and some other isotopes such as the beryllium isotope ¹⁰Be, are formed from the bombardment of the atmosphere by cosmic-ray particles. The ¹⁴C in the atmosphere is converted to ¹⁴CO₂ and incorporated into the tree rings as they form; the year of growth can be precisely determined from dendrochronology. Production of cosmogenic isotopes is high during periods of low solar magnetic activity. But during the Sun's active phase (with

Molecular motors

Smooth coupling in *Salmonella*

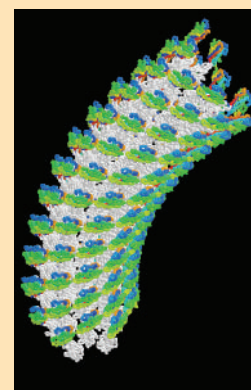
Bacteria such as *Salmonella typhimurium* move by the action of their flagella. Depending on the direction of rotation, flagella either act singly, causing uncoordinated tumbling, or clump together into a single helical propeller for straight-line swimming. The 60-nm-long hook that joins the flagellar filament to its motor in the bacterial cell wall must thus bend through as much as 90° in a millisecond or less, all the time rotating at up to 300 revolutions per second. Elsewhere in this issue (*Nature* **431**, 1062–1068; 2004), Fadel A. Samatey *et al.* describe how they determined the atomic structure of this super-flexible universal joint, and thereby how it achieves such a feat of engineering.

The hook is a hollow tube assembled from 11 chains, or

protofilaments, of a single protein, called FlgE. These protofilaments are stacked together with a slight helical twist that changes slightly with the direction of rotation.

Samatey *et al.* made their model by first determining the structure of the central region of FlgE by X-ray crystallography at a resolution of 1.8 Å. This was then fitted into the lower-resolution images of isolated, straight hooks as seen by electron microscopy. The final curved hook (shown here with individual protein chains coloured from blue through to red) emerged by computer simulation of the squashing and stretching of individual protofilaments.

This modelling showed that the hook's mechanical properties result from a combination of flexibility and



rigidity at the molecular level. Individual protofilaments can grow and shrink in length by as much as 50% through flexing of a hinge that joins the two major domains of FlgE. However, the interlocking of adjacent subunits prevents the protofilaments from sliding against each other. The flagellar hook can thus bend but not twist, allowing efficient transmission of force from motor to propeller. Christopher Surridge

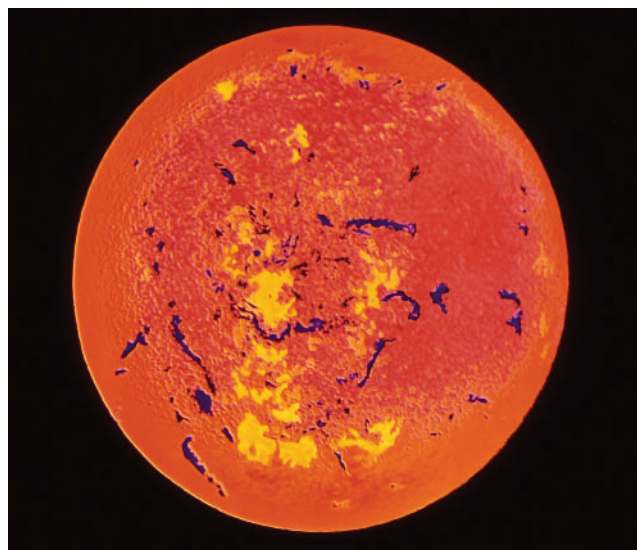


Figure 1 Solar blemishes. This is a false-colour image of the face of the Sun, with sunspots as black patches. In the work discussed here, Solanki *et al.*¹ have produced a reconstruction of sunspot number going back 11,000 years.

high sunspot number), the more intense solar wind — the ions streaming out from the Sun — deflects charged particles so that fewer of them enter Earth's atmosphere.

Solanki and colleagues' first step was to determine the rate of ^{14}C production using the tree-ring record of atmospheric ^{14}C concentration after removing the long-term trend in Earth's magnetic field, which modulates the cosmic-ray flux. The concentrations of ^{14}C in the atmosphere may also be affected by variations in ocean circulation, because carbon is partitioned between the atmosphere, the ocean and the biosphere. But there is no evidence of major oceanic variability over the past 11,000 years, and carbon fluxes in the biosphere are not sufficient to cause large changes in atmospheric ^{14}C .

The second step was to calculate the cosmic-ray flux from the data for ^{14}C production, by 'inverting' a model of the transport and modulation of galactic cosmic rays within the envelope of the solar wind; model inversion means working backwards from the answer to find the necessary input to produce that answer. Solanki *et al.* then reconstructed the Sun's open magnetic flux — the magnetic field that extends into the interplanetary medium — from a model of the effect of the open magnetic flux on the transport of galactic cosmic rays. Finally, a model describing the evolution of the open magnetic flux for a given sunspot number was inverted to produce estimates of sunspot number. Within well-defined limits of uncertainty, the series of models reproduce the observed record of sunspots extremely well, from almost no sunspots during the seventeenth century to the current high levels.

Climate variability on centennial to millennial timescales is documented in many palaeoclimate records going back at least as far as the end of the last glaciation, some 12,000 years ago. Whether solar activity is a dominant influence in these changes is a subject of intense debate^{3–6}. The exact relationship of solar irradiance to sunspot number is

still uncertain^{7,8}, but the reconstructed sunspot number will nonetheless provide a much-needed record of solar activity. This can then be compared with palaeoclimate data sets to test theories of possible solar-climate connections, as well as enabling physicists to model long-term solar variability. A better understanding of the mechanisms responsible for past climate variability will also help those using global circulation models to predict future climate change.

So does the current episode of high sunspot number imply that the Sun has had a significant role in the global warming of the late twentieth century? The answer is no.

Evolutionary biology

Mortality and lifespan

Peter A. Abrams

How does natural selection affect lifespan? The question has exercised biologists for some years. The latest twist comes from ingenious experiments on tropical fish from different ecological backgrounds.

On page 1095 of this issue, Reznick *et al.*¹ describe how they have investigated one of the main factors that influence the evolution of an organism's lifespan. That factor is the risk of dying that a population faces as a result of environmental conditions (such as, in this case, predation). The study subjects are guppies, small tropical fish that are widely used in evolutionary studies, and the authors provide the first experimental support for the prediction that a higher environmental risk of mortality can select for inherently longer-lived organisms.

Guppies from the lower reaches of several rivers in Trinidad are subject to much higher rates of predation than those in the upper parts of the same rivers, where waterfalls block access by larger fish. In predator-free lab experiments, Reznick *et al.* found that guppies from the high-predation segments

Although climate models differ in their estimation of the Sun's contribution to recent warming, even those that include spectrally varying changes in solar irradiance conclude that anthropogenic causes are the prime factor^{9–12}. The high probability that this episode will end soon is not likely to cut us much slack in controlling global warming unless we reduce greenhouse-gas emissions. But because the solar influence may be more regionally variable than the effects of greenhouse gases¹¹, model-based predictions of regional climate change may be improved by this study. It is at the regional level that climate change will have the greatest impact on society. ■

Paula J. Reimer is at the ¹⁴CHRONO Centre for Climate, the Environment and Chronology, Queen's University Belfast, Belfast BT9 6AX, UK. e-mail: p.j.reimer@qub.ac.uk

1. Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M. & Beer, J. *Nature* **431**, 1084–1087 (2004).
2. Stuiver, M. *et al.* *Radiocarbon* **40**, 1041–1083 (1998).
3. Bond, G. *et al.* *Science* **294**, 2130–2136 (2001).
4. Andrews, J. T. *et al.* *Earth Planet. Sci. Lett.* **210**, 453–465 (2003).
5. Blaauw, M., van Geel, B. & van der Plicht, J. *Holocene* **14**, 35–44 (2004).
6. Nesje, A., Dahl, S. O. & Bakke, J. *Holocene* **14**, 299–310 (2004).
7. de Toma, G., White, O. R., Chapman, G. A. & Walton, S. R. *Adv. Space Res.* **34**, 237–242 (2004).
8. Fontenla, J. M. *et al.* *Astrophys. J.* **605**, L85–L88 (2004).
9. Meehl, G. A., Washington, W. M., Wigley, T. M. L., Arblaster, J. M. & Dai, A. *J. Clim.* **16**, 426–444 (2003).
10. Solanki, S. K. & Krivova, N. A. *J. Geophys. Res.* **108**, doi:10.1029/2002JA009753 (2003).
11. Stott, P. A., Jones, G. S. & Mitchell, J. F. B. *J. Clim.* **16**, 4079–4093 (2003).
12. Rind, D., Shindell, D., Perlwitz, J. & Lerner, J. *J. Clim.* **17**, 906–929 (2004).

of two of the rivers lived up to 35% longer than those from low-predation segments of the same watercourse. In addition, the guppies from high-predation sites had a 40% longer reproductive span and reproduced at a higher rate. So a background of higher mortality under natural conditions has apparently led to the evolution of both a longer lifespan and a longer reproductive span. Their longest-lived fish, a female, is pictured in Figure 1.

A bit of history is required to see why this observation is surprising. Environmentally caused ('extrinsic') mortality has long been recognized as a key factor determining how natural selection moulds 'intrinsic' mortality — the death rate that a population would have under some standardized, generally benign, set of environmental conditions. Although evolution should favour lower

intrinsic mortality (and a longer intrinsic lifespan) when all else is equal, many organisms face a trade-off between higher levels of reproduction or lower levels of intrinsic mortality. One of the main reasons that senescence occurs is because repair is costly: resources that are devoted to maintaining an organism are not available for reproduction. In the 1950s, Peter Medawar² and George Williams³ pointed out that high extrinsic mortality could favour shorter intrinsic lifespan. Why, they reasoned, should an organism invest in costly repair that will probably only ensure that it is in prime physical condition when its life ends? Higher extrinsic mortality should favour low investment in repair, and thus a high intrinsic mortality and a short intrinsic lifespan.

But this reasoning didn't take account of two further factors. One is that higher extrinsic mortality also slows the rate of population growth, and more slowly growing populations are expected to evolve to have lower rates of intrinsic mortality and a longer lifespan^{4,5}. The other is the interaction between extrinsic mortality factors and physiological repair or maintenance^{5,6}. If predators can be evaded by fast, but not by slow prey, greater predation risk should select for greater maintenance of the body systems essential for fast movement. This higher level of repair would then prolong intrinsic lifespan.

Higher extrinsic mortality (more predators) could also have indirect effects that Medawar and Williams did not consider. For example, it reduces population size, which in turn increases the abundance of food or other resources. These changes may have their own effects on both population growth and the level of intrinsic mortality favoured by selection. Other complications arise if the mortality factor has a greater effect on some ages than on others^{5,6} — if, for example, predators prefer to capture larger, older prey. As a result of these complicating features, many types of mortality are expected to reduce intrinsic death rates at some ages while increasing them at others⁶. In any event, theory suggests that higher extrinsic mortality will produce evolutionary conditions that can either extend or shorten the intrinsic lifespan.

Given these complexities, the curious feature of previous observational⁷ and experimental work⁸ has been its support for the Medawar–Williams prediction. There have been exceptions⁹, if only suggestively so. But the almost unanimous evidence that high extrinsic mortality is associated with shorter lifespan is puzzling because there is no reason to believe that the conditions that produce the opposite outcome are rare in nature.

So it is reassuring that Reznick *et al.*¹ found longer intrinsic lifespans in guppies from populations characterized by higher predation rates. The authors also looked at whether these evolutionary changes might



Figure 1 Star survivor — the longest lived of the guppies studied by Reznick *et al.*¹. The photo was taken shortly before her death at the age of 1,464 days.

be an indirect consequence of predator-caused deaths, such as the availability, in natural settings, of more food for the remaining guppies. Reznick and colleagues' study is unique in examining this effect. They found that food alone could not account for the difference in intrinsic mortalities seen in their experiments, but that having more food enhanced the lifespan-lengthening effect of a high-predation background.

There is no doubt that the guppies from high-predation sites have both longer intrinsic lifespans and longer reproductive spans. But is it valid to conclude that they have slower senescence? This is a more difficult question. A hypothetical population with no senescence (that is, no age-related decline in survival or reproduction) could still have a short lifespan if it had a high mortality rate that was independent of age. If guppies from high-predation sites begin their adult life with a lower rate of intrinsic mortality than those from low-predation environments, they could have the same rate of increase with age in their mortality rate, but would still have a longer lifespan. One could then argue that the two populations had identical rates of senescence. Some measures of the rate of change of intrinsic mortality with age suggest that senescence is delayed in guppies from high-predation sites.

However, senescence encompasses relationships between many different components of fitness and age, none of which can be adequately summarized by a single number: there are many potential measures of the rate of senescence, and conclusions about this rate depend on the measure chosen. It might be possible to make a case that guppies from high-predation environments are more robust, but age at a rate equal to or higher than that of low-predation guppies. Regardless of the mathematical measure used to quantify the rate of senescence, the work of Reznick *et al.* clearly shows that rates of

senescence differ among the different components of fitness examined: survival, reproduction or swimming performance. The reasons for these differences are not yet understood.

It would be surprising if guppies were the only species for which an added risk of mortality lengthens intrinsic lifespan. Similar studies on other species will help us understand the underlying reasons why Medawar and Williams' predictions hold for some species and not for others. Such studies should follow Reznick and colleagues' lead in quantifying declines in several fitness components and studying the indirect ecological consequences of higher extrinsic mortality.

Peter A. Abrams is in the Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 3G5, Canada.

e-mail: abrams@zoo.utoronto.ca

1. Reznick, D. N., Bryant, M. J., Roff, D., Ghalambor, C. K. & Ghalambor, D. E. *Nature* **431**, 1095–1099 (2004).
2. Medawar, P. B. *An Unsolved Problem in Biology* (Lewis, London, 1952).
3. Williams, G. C. *Evolution* **11**, 398–411 (1957).
4. Charlesworth, B. A. *Evolution in Age-Structured Populations* (Cambridge Univ. Press, 1980).
5. Abrams, P. A. *Evolution* **47**, 877–887 (1993).
6. Williams, P. D. & Day, T. *Evolution* **57**, 1478–1488 (2002).
7. Ricklefs, R. E. *Am. Nat.* **152**, 24–44 (1998).
8. Stearns, S. C., Ackermann, M., Doebe, M. & Kaiser, M. *Proc. Natl Acad. Sci. USA* **97**, 3309–3313 (2000).
9. Miller, R. A., Harper, J. M., Dysko, R. C., Durkee, S. J. & Austad, S. N. *Exp. Biol. Med.* **227**, 500–508 (2002).

Correction

In Yi Zhang's News and Views article "Molecular biology: No exception to reversibility" (*Nature* **431**, 637–639; 2004), there were errors in Fig. 1b. In the side chain of citrulline, a double bond should have been shown between NH and O, rather than between NH and NH₂. Several of the connecting atoms are erroneously shown as H rather than N. And the leaving methylamine should have been represented as +NH₃CH₃ rather than +NH₂CH₃.

Neuroscience

Depression and gene expression

Proc. Natl Acad. Sci. USA **101**, 15506–15511 (2004)

Major depressive disorder is a serious affliction for all too many people. Using microarray technology, S. J. Evans *et al.* show that changes in certain gene transcripts could help to explain the molecular basis of this condition.

The authors assessed levels of gene activity in post-mortem samples of brain tissue from patients who had suffered from major depressive disorder, as well as another mood disorder, and from control subjects with no history of psychiatric problems. They found altered expression of fibroblast growth factor (FGF) in the depressed group only; in fact, in people with major depressive disorder, several genes encoding FGFs were coordinately disrupted. These findings support the hypothesis that altered growth-factor activity in the brain contributes to mood disorders.

Evans *et al.* also demonstrate that certain antidepressants might have a moderating effect on changes in the levels of some FGF transcripts: these changes were less profound in patients taking drugs known as 'specific serotonin reuptake inhibitors'. The researchers accordingly point out that the FGF system could be involved in the mechanism by which antidepressants operate.

Roxanne Khamis

Astrophysics

Ageing dwarfs still active

Astrophys. J. Rapid release 6 October 2004

The hot, ionized gas swirling inside a star generates a magnetic field. This dynamo effect is characteristic of all but the coolest stars, and in the star's outer atmosphere, or corona, it can generate flares of material that emit a sudden burst of X-rays.

But brown dwarfs are different. These glowing cinders are 'failed' stars that do not have enough mass to generate the pressure needed for hydrogen fusion. Some astronomers believe that the dynamo switches off in older, cooler dwarfs, as their ionized material becomes neutral.

B. Stelzer now reports an X-ray flare from an older brown dwarf (Gl 569Bab), and sustained X-ray emission from its corona. She claims that this is probably the first detection of quiescent X-ray emission from a brown-dwarf corona, although she notes the possibility that the X-rays are some kind of afterglow from the flare.

Only one previous example of an X-ray flare from an older brown dwarf is known. This new sighting not only confirms that even ageing brown dwarfs

Animal behaviour

Tools and termites

Am. Nat. **164**, 567–581 (2004)

Tool use by apes is well documented. But according to Crickette Sanz and colleagues, who have followed the behaviour of groups of chimpanzees in the Nouabalé-Ndoki National Park in Congo, there is more to be discovered.

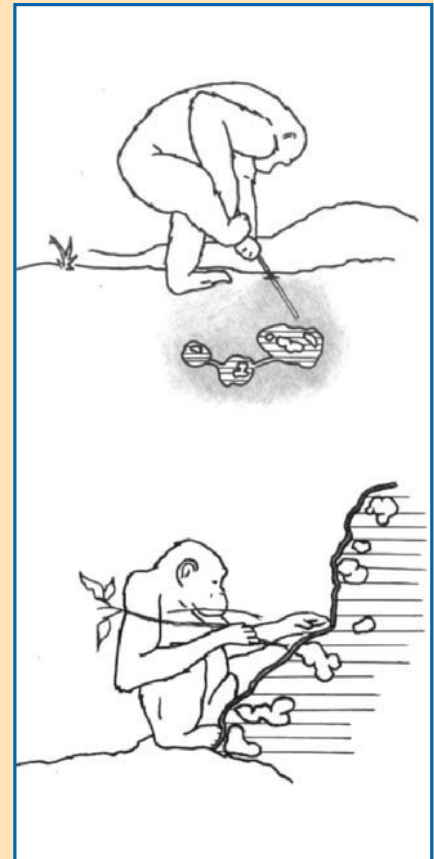
Over six months, Sanz *et al.* used remote video recording to observe how chimps tackled the tempting meals of termites to be found in nests both below and above ground.

The lower drawing shows an adult female, deep in concentration at a termite mound, and summarizes one of the authors' main findings.

They found that chimps used one kind of tool, a 'perforating twig', to open up sealed holes on the mound surface. As the termites rush to defend their home, the chimp switches to a more delicate 'brush-tip fishing probe', seen here in her mouth, to extract the termite dinner. Underground nests were breached with a different implement, a 'puncturing stick', as shown in the upper sketch. Subtle differences in practice between chimp groups raise the question as to what extent ecological factors or different group traditions are responsible.

Tim Lincoln

Video recordings of chimpanzee performance are at www.journals.uchicago.edu/AN/journal/issues/v164n5/40471/40471.html



D. MORGAN, WILDLIFE CONSERV. SOC.

can sustain an active magnetic field, but also that they can produce a constant X-ray glow from a heated corona even in their chilly later years.

Mark Peplow

Mechanics

On the rebound

Phys. Rev. Lett. **93**, 154301 (2004)

How can a ball bounce up from the floor at a steeper angle than that at which it struck it? Hiroto Kuninaka and Hisao Hayakawa show that this seemingly bizarre phenomenon has a perfectly rational explanation.

Their calculations and computer simulations show that it can happen when a hard ball hits a softer surface. Deformation of the surface — which is pushed into a little ridge behind the point of impact — causes some of the ball's horizontal velocity to be transferred to vertical velocity. This can produce a coefficient of normal restitution, e — in effect, the ratio of the final to initial vertical components of the velocity — that is greater than 1. Kuninaka and Hayakawa predict that for sufficiently oblique impacts (about 11°), e can be as large as 1.3.

These findings bear out earlier empirical observations showing that e could exceed 1 for glancing impacts. But as the wall must be softer than the ball, it won't work on the squash court.

Philip Ball

Population biology

Cannibal cycling

Proc. R. Soc. Lond. B (suppl.)
doi:10.1098/rsbl.2004.0231 (2004)

The curious fluctuations in populations of the Indonesian 28-spot ladybeetle, *Epilachna vigintioctopunctata*, may have a rather grisly explanation. The cycles have a period roughly the same as the insect's development time to adulthood, and are only seen in juvenile stages. They are best explained, say Koji Nakamura and colleagues, by the larvae turning cannibal when food is short.

If the pattern were due to a natural enemy, such as a predator or parasite, both adult and juvenile populations of the ladybeetle would fluctuate. So Nakamura *et al.* propose that the cycles arise from within the beetle species — as larval populations grow, the larvae are forced to start eating the eggs of their fellow ladybeetles.

The authors examined data from *E. vigintioctopunctata* censuses taken in Indonesia in the 1980s and 1990s. At Padang, the beetles showed a generation time of 45 days, and their egg numbers rose and fell roughly every 50 days. This provides real-world support for the cannibalism theory, and mirrors earlier findings in lab cultures of the flour beetle *Tribolium*.

Michael Hopkin

English elm is a 2,000-year-old Roman clone

This tree's genetic uniformity may have helped to fell entire European populations.

S. DOMÍNGUEZ

The outbreak of Dutch elm disease in the 1970s ravaged European elm populations, killing more than 25 million trees in Britain alone; the greatest impact was on *Ulmus procera*, otherwise known as the English elm¹. Here we use molecular and historical information to show that this elm derives from a single clone that the Romans transported from Italy to the Iberian peninsula, and from there to Britain, for the purpose of supporting and training vines. Its highly efficient vegetative reproduction and its inability to set seeds have preserved this clone unaltered for 2,000 years as the core of the English elm population — and the preponderance of this susceptible variety may have favoured a rapid spread of the disease.

Following the results from the European Union project RESGEN CT96-78 on elm genetic resources, we studied the variability of chloroplast DNA in the two elm species native to Spain (the field elm, *U. minor* Miller *sensu latissimo*², and the wych elm, *U. glabra* Huds.) and in elm samples from France, Greece, Italy and Britain (for methods, see supplementary information).

We analysed restriction fragment-length polymorphisms in chloroplast DNA after amplification by polymerase chain reaction and detected four major chloroplast lineages, which are maternally inherited in *Ulmus* (see supplementary information). Lineage A appears in Greece; lineage B in northern Italy; lineage C in Italy, the Iberian peninsula and Britain (but not in France); and lineage D throughout all of Spain, northern Italy, France and Britain. Lineage-D haplotypes appear in both *U. minor* and *U. glabra* samples. By contrast, in Spain and Britain, lineage-C haplotypes are observed only in *U. minor*, although they appear in both species in Italy (mainly in elms from the central region).

Except for the presence of lineage C in the Iberian peninsula and Britain, the distribution of lineages is very similar to those described for other trees³. Haplotype sharing among tree species is reckoned to indicate their presence in common glacial refugia, where chloroplasts would have been exchanged through hybridization and introgression⁴. Absence of lineage-C haplotypes in Spanish *U. glabra* highlights the lack of refugia for this lineage in Iberia, whereas its presence in both elms in Italy indicates that this lineage has an Italian origin.

Eighteen field elms with the same haplotype (from lineage C), collected from across Spain, Britain and central Italy, as well as five lineage-D individuals from Spain and Britain, were genotyped using seven nuclear



microsatellites⁵ and two amplified fragment-length polymorphism (AFLP) primer combinations. The results revealed a widely distributed clone within lineage-C elms, represented here by five Spanish and three British samples (see supplementary information). We detected slight differences among the AFLP fingerprinting profiles of these eight samples, attributable to somatic mutations. The closest genetic similarity was observed between the clone and samples from Latium in Italy, which reinforces the idea that its origin was Roman.

These eight individuals were classified as English elm (*U. minor* var. *vulgaris*², commonly called *U. procera* Salisb. by British botanists²). In Britain² and in Spain⁶, English elm rarely sets seeds, but produces pollen normally and is very effective in vegetative propagation. These reproductive features, together with the deliberate plantation of this variety of elm by humans, could explain the maintenance of this genotype and its spread over Spain and Britain. But when, whence and why was it transported?

Although it has been suggested that the English elm was introduced during the Bronze Age by Celtic tribes², our results support a hypothesis⁷ that it corresponds to the Atinian elm, which was used for vine-training by the Romans. In his treaty *De Re Rustica*⁸ (written in about AD 50), the Roman agronomist Columella advocates the use of elm for this purpose, recommending in particular a barren tree that was vegetatively propagated — the Atinian elm.

Columella owned three farms in Latium⁹, where most Italian lineage-C samples are found, and a fourth vineyard in Xerez¹⁰

(Andalusia, Spain). He and other farmers may have introduced different Italian elms to the Iberian peninsula, including the Atinian elm. Columella's writings influenced the subsequent establishment of vineyards to such an extent that the Roman emperor Domitian prohibited the plantation of new vines in Italy in AD 92 and ordered half of the vineyards in the provinces to be cut down¹¹. In the meantime, the Atinian clone spread across Iberia and was probably transported to Britain in the form of root suckers, as indicated by the presence of elm pollen in a Roman vineyard in Britain¹² and by the coincident distribution of suspected Roman vineyards¹² and *U. procera* in Britain².

The identification of the English elm with the Atinian clone was first proposed in the nineteenth century⁷. Our findings provide molecular support for this proposal and indicate that the English elm originated from the massive propagation of the Atinian clone by the Romans. This large-scale transformation of the elm's natural diversity became critical in the twentieth century, when most English elms succumbed to Dutch elm disease¹, and should be taken into account in current European elm breeding and conservation strategies.

Luis Gil*, Pablo Fuentes-Utrilla*, Álvaro Soto†, M. Teresa Cervera†, Carmen Collada‡

*Departamento de Silvopascicultura, and

‡Departamento de Biotecnología, Escuela Técnica Superior de Ingenieros de Montes (UPM), Ciudad Universitaria s/n, 28040 Madrid, Spain
e-mail: lgil@montes.upm.es

†Instituto Nacional de Investigaciones Agrarias, Carretera de La Coruña, 28040 Madrid, Spain

1. Brasier, C. in *Forest Research* 20–28 (Forestry Commission Report, Edinburgh, 1996).

2. Richens, R. H. *Elm* (Cambridge Univ. Press, Cambridge, 1983).

3. Hewitt, G. M. *Biol. J. Linn. Soc.* **68**, 87–112 (1999).

4. Dumolin-Lapègue, S., Demesure, B., Fineschi, S., Le Corre, V. & Petit, R. J. *Genetics* **146**, 1475–1487 (1997).

5. Collada, C., Fuentes-Utrilla, P., Gil, L. & Cervera, M. T. *Mol. Ecol. Notes* (in the press).

6. López-Almansa, J. C., Pannell, J. R. & Gil, L. *Am. J. Bot.* **90**, 603–609 (2003).

7. Heybroek, H. M. *Forestry Compendium CD-ROM* (CAB International, 2000).

8. Forster, E. S. & Heffner, E. (translators) *Lucius Junius Moderatus Columella: On Agriculture II* (Harvard Univ. Press, Cambridge, Massachusetts, 1954; reprinted 1977).

9. Ash, H. B. (translator) *Lucius Junius Moderatus Columella: On Agriculture I* (Harvard Univ. Press, Cambridge, Massachusetts, 1941; reprinted 1977).

10. Tovar, A. in *Homenaje Nacional a Lucio Junio Moderato Columela* 93–99 (Asociación de Publicistas y Escritores Agrarios Españoles, Cádiz, 1975).

11. Martin, R. *Recherches sur les Agronomes Latins et leurs Conceptions Economiques et Sociales* (Les Belles Lettres, Paris, 1971).

12. Brown, A. G., Meadows, I., Turner, S. D. & Mattingly, D. J. *Antiquity* **75**, 745–757 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia

P. Brown¹, T. Sutikna², M. J. Morwood¹, R. P. Soejono², Jatmiko², E. Wayhu Saptomo² & Rokus Awe Due²

¹Archaeology & Palaeoanthropology, School of Human & Environmental Studies, University of New England, Armidale, New South Wales 2351, Australia

²Indonesian Centre for Archaeology, Jl. Raya Condet Pejaten No. 4, Jakarta 12001, Indonesia

Currently, it is widely accepted that only one hominin genus, *Homo*, was present in Pleistocene Asia, represented by two species, *Homo erectus* and *Homo sapiens*. Both species are characterized by greater brain size, increased body height and smaller teeth relative to Pliocene *Australopithecus* in Africa. Here we report the discovery, from the Late Pleistocene of Flores, Indonesia, of an adult hominin with stature and endocranial volume approximating 1 m and 380 cm³, respectively—equal to the smallest-known australopithecines. The combination of primitive and derived features assigns this hominin to a new species, *Homo floresiensis*. The most likely explanation for its existence on Flores is long-term isolation, with subsequent endemic dwarfing, of an ancestral *H. erectus* population. Importantly, *H. floresiensis* shows that the genus *Homo* is morphologically more varied and flexible in its adaptive responses than previously thought.

The LB1 skeleton was recovered in September 2003 during archaeological excavation at Liang Bua, Flores¹. Most of the skeletal elements for LB1 were found in a small area, approximately 500 cm², with parts of the skeleton still articulated and the tibiae flexed under the femora. Orientation of the skeleton in relation to site stratigraphy suggests that the body had moved slightly down slope before being covered with sediment. The skeleton is extremely fragile and not fossilized or covered with calcium carbonate. Recovered elements include a fairly complete cranium and mandible, right leg and left innominate. Bones of the left leg, hands and feet are less complete, while the vertebral column, sacrum, scapulae, clavicles and ribs are only represented by fragments. The position of the skeleton suggests that the arms are still in the wall of the excavation, and may be recovered in the future. Tooth eruption, epiphyseal union and tooth wear indicate an adult, and pelvic anatomy strongly supports the skeleton being that of a female. On the basis of its unique combination of primitive and derived features we assign this skeleton to a new species, *Homo floresiensis*.

Description of *Homo floresiensis*

Order Primates Linnaeus, 1758
Suborder Anthropoidea Mivart, 1864
Superfamily Hominoidea Gray, 1825
Family Hominidae Gray, 1825
Tribe Hominini Gray, 1825
Genus *Homo* Linnaeus, 1758
Homo floresiensis sp. nov.

Etymology. Recognizing that this species has only been identified on the island of Flores, and a prolonged period of isolation may have resulted in the evolution of an island endemic form.

Holotype. LB1 partial adult skeleton excavated in September 2003. Recovered skeletal elements include the cranium and mandible, femora, tibiae, fibulae and patellae, partial pelvis, incomplete hands and feet, and fragments of vertebrae, sacrum, ribs, scapulae and clavicles. The repository is the Centre for Archaeology, Jakarta, Indonesia.

Referred material. LB2 isolated left mandibular P₃. The repository is the Centre for Archaeology, Jakarta, Indonesia.

Localities. Liang Bua is a limestone cave on Flores, in eastern Indonesia. The cave is located 14 km north of Ruteng, the provincial capital of Manggarai Province, at an altitude of 500 m above sea level and 25 km from the north coast. It occurs at the base of a limestone

hill, on the southern edge of the Wae Racang river valley. The type locality is at 08° 31' 50.4" south latitude 120° 26' 36.9" east longitude.

Horizon. The type specimen LB1 was found at a depth of 5.9 m in Sector VII of the excavation at Liang Bua. It is associated with calibrated accelerator mass spectrometry (AMS) dates of approximately 18 kyr and bracketed by luminescence dates of 35 ± 4 kyr and 14 ± 2 kyr. The referred isolated left P₃ (LB2) was recovered just below a disconformity at 4.7 m in Sector IV, and bracketed by a U-series date of 37.7 ± 0.2 kyr on flowstone, and 20 cm above an electron-spin resonance (ESR)/U-series date of 74⁺¹⁴₋₁₂ kyr on a *Stegodon* molar.

Diagnosis. Small-bodied bipedal hominin with endocranial volume and stature (body height) similar to, or smaller than, *Australopithecus afarensis*. Lacks masticatory adaptations present in *Australopithecus* and *Paranthropus*, with substantially reduced facial height and prognathism, smaller postcanine teeth, and posteriorly orientated infraorbital region. Cranial base flexed. Prominent maxillary canine juga form prominent pillars, laterally separated from nasal aperture. Petrous pyramid smooth, tubular and with low relief, styloid process absent, and without vaginal crest. Superior cranial vault bone thicker than *Australopithecus* and similar to *H. erectus* and *H. sapiens*. Supraorbital torus arches over each orbit and does not form a flat bar as in Javan *H. erectus*. Mandibular P₃ with relatively large occlusal surface area, with prominent protoconid and broad talonid, and either bifurcated roots or a mesiodistally compressed Tomes root. Mandibular P₄ also with Tomes root. First and second molar teeth of similar size. Mandibular coronoid process higher than condyle, and the ramus has a posterior orientation. Mandibular symphysis without chin and with a posterior inclination of the symphyseal axis. Posteriorly inclined alveolar planum with superior and inferior transverse tori. Ilium with marked lateral flare. Femur neck long relative to head diameter, the shaft circular and without pilaster, and there is a high bicondylar angle. Long axis of tibia curved and the midshaft has an oval cross-section.

Description and comparison of the cranial and postcranial elements

Apart from the right zygomatic arch, the cranium is free of substantial distortion (Figs 1 and 2). Unfortunately, the bregmatic region, right frontal, supraorbital, nasal and subnasal regions were damaged when the skeleton was discovered. To repair post-mortem

pressure cracks, and stabilize the vault, the calvarium was dismantled and cleaned endocranially before reconstruction. With the exception of the squamous suture, most of the cranial vault sutures are difficult to locate and this problem persists in computed tomography (CT) scans. As a result it is not possible to locate most of the standard craniometric landmarks with great precision.

The LB1 cranial vault is long and low. In comparison with adult *H. erectus* (including specimens referred to as *Homo ergaster* and *Homo georgicus*) and *H. sapiens* the calvarium of LB1 is extremely small. Indices of cranial shape closely follow the pattern in *H. erectus* (Supplementary Table 1). For instance, maximum cranial breadth is in the inflated supramastoid region, and the vault is broad relative

to its height. In posterior view the parietal contour is similar to *H. erectus* but with reduced cranial height^{2,3}. Internal examination of the neurocranium, directly and with CT scan data, indicates that the brain of LB1 had a flattened platycephalic shape, with greatest breadth across the temporal lobes and reduced parietal lobe development compared with *H. sapiens*. The cranial base angle (basion–sella–foramen caecum) of 130° is relatively flexed in comparison with both *H. sapiens* (mean 137°–138° (refs 4, 5)) and Indonesian *H. erectus* (Sambungmacan 4 141° (ref. 6)). Other small-brained hominins, for instance STS 5 *Australopithecus africanus*, have the primitive less-flexed condition.

The endocranial volume, measured with mustard seed, is



Figure 1 The LB1 cranium and mandible in lateral and three-quarter views, and cranium in frontal, posterior, superior and inferior views. Scale bar, 1 cm.

380 cm³, well below the previously accepted range for the genus *Homo*⁷ and equal to the minimum estimates for *Australopithecus*⁸. The endocranial volume, relative to an indicator of body height (maximum femur length 280 mm), is outside the recorded hominin normal range (Fig. 3). Medially, laterally and basally, the cranial vault bone is thick and lies within the range of *H. erectus* and *H. sapiens*^{9,10} (Supplementary Table 1 and Fig. 2). Reconstruction of the cranial vault, and CT scans, indicated that for most of the cranial vault the relative thickness of the tabular bone and diploë are similar to the normal range in *H. erectus* and *H. sapiens*. In common with *H. erectus* the vault in LB1 is relatively thickened posteriorly and in areas of pneumatization in the lateral cranial base. Thickened vault bone in LB1, relative to that in *Australopithecus* and early *Homo*², results in a substantially reduced endocranial volume in comparison to Plio-Pleistocene hominins with similar external vault dimensions.

The occipital of LB1 is strongly flexed, with an occipital curvature angle of 101° (Supplementary Information), and the length of the nuchal plane dominates over the occipital segment. The occipital torus forms a low extended mound, the occipital protuberance is not particularly prominent compared with Indonesian *H. erectus* and there is a shallow supratatorial sulcus. The endinion is positioned 12 mm inferior to the inion, which is within the range of *H. erectus* and *Australopithecus*¹⁰. Compared with *Australopithecus* and early *Homo*² the foramen magnum is narrow (21 mm) relative to its length (28 mm), and mastoid processes are thickened medio-laterally and are relatively deep (20.5 mm). In common with Asian, and some African, *H. erectus* a deep fissure separates the mastoid process from the petrous crest of the tympanic^{10,11}. Bilaterally there is a recess between the tympanic plate and the entoglenoid pyramid. These two traits are not seen in modern humans, and show varied levels of development in Asian and African *H. erectus* and Pliocene hominins¹⁰. The depth and breadth of the glenoid fossae and angulation of the articular eminence are within the range of variation in *H. sapiens*. The inferior surface of the petrous pyramid has numerous similarities with Zhoukoudian *H. erectus*¹², with a smooth tubular external surface as in chimpanzees, and a constricted foramen lacerum. Styloid processes and vaginal crests are not present.

The temporal lines approach to within 33 mm of the coronal suture and have a marked posterior extension. There are no raised angular tori as is common in *H. erectus*¹⁰ and some terminal Pleistocene Australians, and no evidence of parietal keeling. Posteriorly there is some asymmetrical obelionic flattening and CT scans

indicate that the parietals reduce in thickness in this slightly depressed area (Fig. 2). A principal component analysis (PCA) of five cranial vault measurements separates LB1, STS5 (*A. africanus*) and KNM-ER 1813 (early *Homo*) from other hominin calvaria in size and shape. Shape, particularly height and breadth relationships, placed LB1 closest to ER-3883, ER-3733 and Sangiran 2 *H. erectus* (Supplementary Fig. 1).

The face of LB1 lacks most of the masticatory adaptations evident in *Australopithecus* and its overall morphology is similar to members of the genus *Homo*^{2,3}. In comparison with *Australopithecus*, tooth dimensions and the alveolar segment of the maxillae are greatly reduced, as are facial height and prognathism. The facial skeleton is dominated by pronounced canine juga, which form prominent pillars lateral to the nasal aperture. However, these are distinct from the anterior pillars adjacent to the nasal aperture in *A. africanus*^{2,3}. The infraorbital fossae are deep with large infra-orbital foramina, the orbits have a particularly arched superior border and a volume of 15.5 cm³ (ref. 13). On the better preserved right-hand side, the supraorbital torus arches over the orbit and does not form a straight bar, with bulbous laterally projecting trigones, as in Indonesian *H. erectus*¹¹. The preserved section of the right torus only extends medially slightly past mid-orbit, and the morphology of the glabella region and medial torus is unknown. In facial view the zygo-maxillary region is medially deep relative to facial height, and the inferior border of the malars are angled at 55° relative to the coronal plane. In lateral view the infraorbital region is orientated posteriorly as in other members of the genus *Homo*, rather than the more vertical orientation in *A. africanus*^{2,3}. The root of the maxillary zygomatic process is centred above the first molar, and the incisive canal is relatively large and has an anterior location, contrasting with African and Javan *H. erectus*. In lateral view, curvature of the frontal squama is more similar to African early *Homo* and Dmanisi *H. ergaster*^{3,14} than it is to the Javan hominins. The frontal squama is separated from the supraorbital torus by a supraorbital sulcus. In the middle third of the frontal there is a slight sagittal keel, extending into the remains of a low, broad prebregmatic eminence. On the midfrontal squama there is a circular healed lesion, probably the remains of a depressed fracture, which is about 15 mm across.

The mandible is complete, apart from some damage to the right condyle (Fig. 4) and combines features present in a variety of Pliocene and Pleistocene hominins. Post-mortem breaks through the corpus at the right P₃ and M₂, and the left canine have resulted in some lateral distortion of the right ramus. There is a strong Curve of

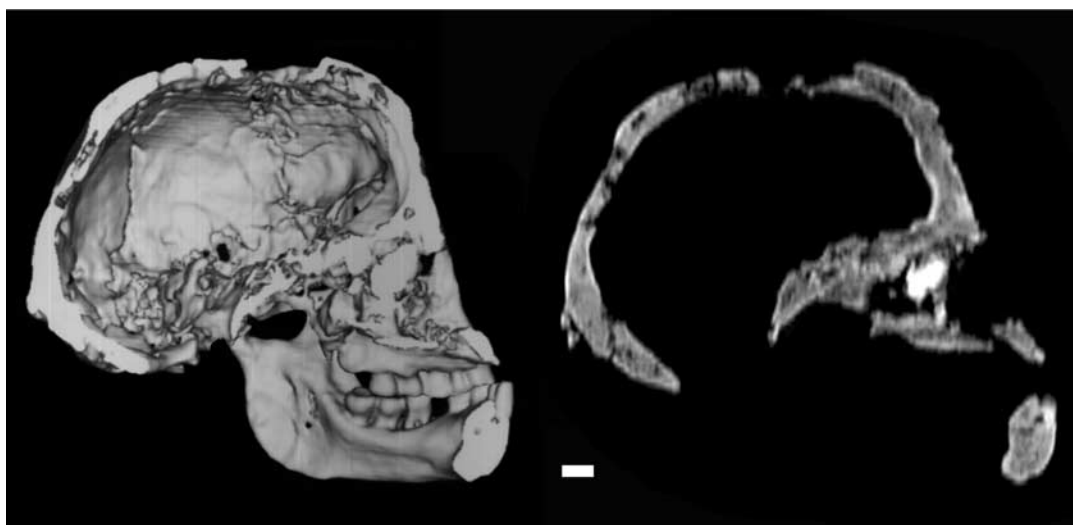


Figure 2 Rendered three-dimensional and individual midsagittal CT section views of the LB1 cranium and mandible. Scale bar, 1 cm.

Spee. The ramus root inserts on the corpus above the lateral prominence, and in lateral aspect obscures the distal M_3 . The ramus is broadest inferiorly, slopes slightly posteriorly and is thickened medio-laterally, and the coronoid process is higher than the condyle. The right condyle has a maximum breadth of 18 mm. There is a narrow and shallow extramolar sulcus and moderate lateral prominence. The anterior portion of the corpus is rounded and bulbous and without a chin. In the posterior symphyseal region the alveolar planum inclines postero-inferiorly, there is a moderate superior torus, deep and broad diagastric fossa, and the inferior transverse torus is low and rounded rather than shelf-like (Fig. 4). There is a strong posterior angulation of the symphyseal axis, and the overall morphology of the symphysis is very similar to LH4 *A. afarensis* and unlike Zhoukoudian and Sangiran *H. erectus*. There are bilaterally double mental foramina, with the posterior foramina smaller and located more inferiorly. Double mental foramina are common in Indonesian *H. erectus*¹⁵. While the mandibular dental arch is narrow anteriorly, and long relative to its breadth, the axis of P_3 – M_3 is laterally convex rather than straight (Fig. 4).

The right P_4 is absent and the alveolus completely fused, the left P_4 was lost after death, and CT scans indicate that the maxillary right M^3 was congenitally absent. The relatively small and conical alveolus for the missing left M^3 suggests that it had a much smaller crown than M^1 and M^2 . Size, spacing and angulation of the maxillary incisor alveoli, and absence of a mesial facet on the canines suggest that incisor I^2 was much smaller than I^1 , and there may have been a diastema. Occlusal wear has removed details of cusp and fissure morphology from most of the maxillary and mandibular teeth. The canines have worn down to a relatively flat surface and there would have been an edge-to-edge bite anteriorly. Interproximal wear is pronounced and in combination with the loss of crown height means that mesio-distal crown dimensions convey little phylogenetic information. With the exception of P_3 the size and morphology of the mandibular teeth follow the pattern in *H. erectus* and *H. sapiens* (Fig. 5, Supplementary Table 2). There is not a great deal of difference between the size of the molar teeth in each quadrant, and the size sequence for both mandibular and maxillary teeth is $M1 \geq M2 > M3$. Using the megadontal quotient as a measure of relative tooth size¹⁶, and substituting P_3 crown area for the missing

P_4 s, LB1 is megadont (1.8) relative to *H. sapiens* (0.9) and *H. ergaster* (0.9), but not *H. habilis* (1.9) (ref. 8) (Supplementary Information). The P_3 s have a relatively great occlusal surface area (molariform) and when unworn had a prominent protoconid and broad talonid. Both P_3 s have bifurcated roots and the alveolus for the left P_4 indicates a mesiodistally compressed, broad Tomes' root. A larger, less worn, isolated left P_3 from the deposit (LB2) has a more triangular occlusal outline, and a Tomes' root (Supplementary Fig. 2). Mandibular P_3 s and P_4 s with similar crown and root morphology have been recorded for *Australopithecus* and early *Homo*^{17,18}, and some Indonesian *H. erectus* mandibular premolars also have bifurcated or Tomes' roots¹⁵. Unusually, both maxillary P_4 s are rotated parallel to the tooth row, a trait that seems to be unrecorded in any other hominin. Maxillary canines and P_3 s have long roots and very prominent juga. The P_3 juga are emphasized by the rotation of the adjacent P_4 roots.

The pelvic girdle is represented by a right innominate, with damage to the iliac crest and pubic region, and fragments of the sacrum and left innominate. The right innominate, which is undistorted, has a broad greater sciatic notch suggesting that LB1 is a female (Fig. 6). In common with all bipedal hominins, the iliac blade is relatively short and wide¹⁹; however, the ischial spine is not particularly pronounced. Compared with modern humans the LB1 ilium has marked lateral flare, and the blade would have projected

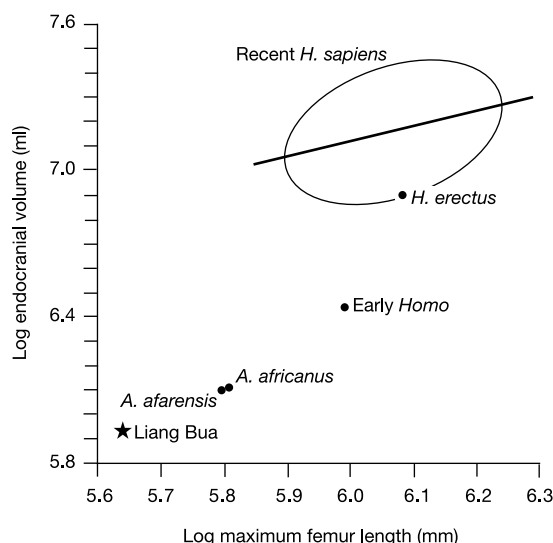


Figure 3 Relationship between endocranial volume and femur length in LB1, *A. afarensis*, *A. africanus*, early *Homo* sp., *H. erectus* and modern *H. sapiens*. Modern human data, with least squares regression line and 95% confidence ellipse, from a global sample of 155 individuals collected by P.B. Details of the hominin samples are in the Supplementary Information.



Figure 4 Right lateral and occlusal views of the LB1 mandible, sagittal profile of the symphysis, occlusal view of the mandibular dentition and occlusal views of the mandibular premolars. Scale bars, 1 cm.

more laterally from the body, relative to the plane of the acetabulum. The left acetabulum is of circular shape, and has a maximum width of 36 mm.

Apart from damage to the lateral condyle and distal shaft, the right femur is complete and undistorted (Fig. 7). The overall anatomy of the femur is most consistent with the broad range of variation in *H. sapiens*, with some departures that may be the result of the allometric effects of very small body size. The femur shaft is

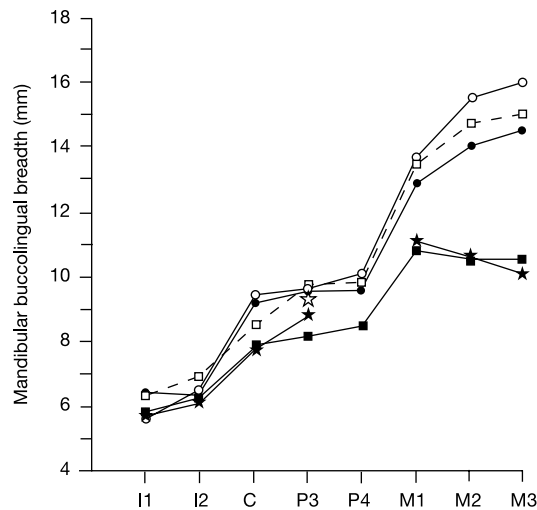


Figure 5 Mean buccolingual tooth crown breadths for mandibular teeth in *A. afarensis* (filled circles), *A. africanus* (open circles), early *Homo* sp. (open squares), modern *H. sapiens* (filled squares), LB1 (filled stars) and LB2 (open stars). There are no mandibular P₄s preserved for LB1. Data for *Australopithecus* and early *Homo* are from ref. 49. Modern human data from a global sample of 1,199 individuals collected by P.B.

relatively straight, and areas of muscle attachment, including the linea aspera, are not well developed. In contrast with some examples of Asian and African *H. erectus*, the femora do not have reduced medullary canals²⁰. On the proximal end, the lesser trochanter is extremely prominent and the strong development of the intertrochanteric crest is similar to *H. sapiens* rather than the flattened intertrochanteric area in *Australopithecus* and *H. erectus* (KNM-ER 1481A, KNM-WT 15000). The biomechanical neck length is 55.5 mm and the neck is long relative to the femoral head diameter (31.5 mm), as is common to both *Australopithecus* and early *Homo*¹⁹. The neck-head junction is 31.5 mm long, with a shaft-neck angle of 130°, and the femur neck is compressed antero-posteriorly (Fig. 7). Several indices of femoral size and shape, for example the relationship between femoral head size and midshaft circumference (66 mm), and femur length and sub-trochanteric shaft size²¹, fall within the chimpanzee and australopithecine range of variation. The femur shaft does not have a pilaster, is circular in cross-section, and has cross-sectional areas of 370 mm² at the midshaft and 359 mm² at the midneck. It is therefore slightly more robust than the best-preserved small-bodied hominin femur of similar length (AL288-1; ref. 21). Distally there is a relatively high bicondylar angle of 14°, which overlaps with that found in *Australopithecus*²².

The right tibia is complete apart from the tip of the medial malleolus (Fig. 7). Its most distinctive feature, apart from its small size (estimated maximum length 235 mm, bicondylar breadth 51.5 mm) and the slight curvature in the long axis, is a shaft that is oval in cross-section (midshaft 347 mm²), without a sharp anterior border, and relatively thickened medio-laterally in the distal half. The relationship between the midshaft circumference and the length of the tibia is in the chimpanzee range of variation and distinct from *Homo*²¹.

Additional evidence of a small-bodied adult hominin is provided by an unassociated left radius shaft, without the articular ends, from

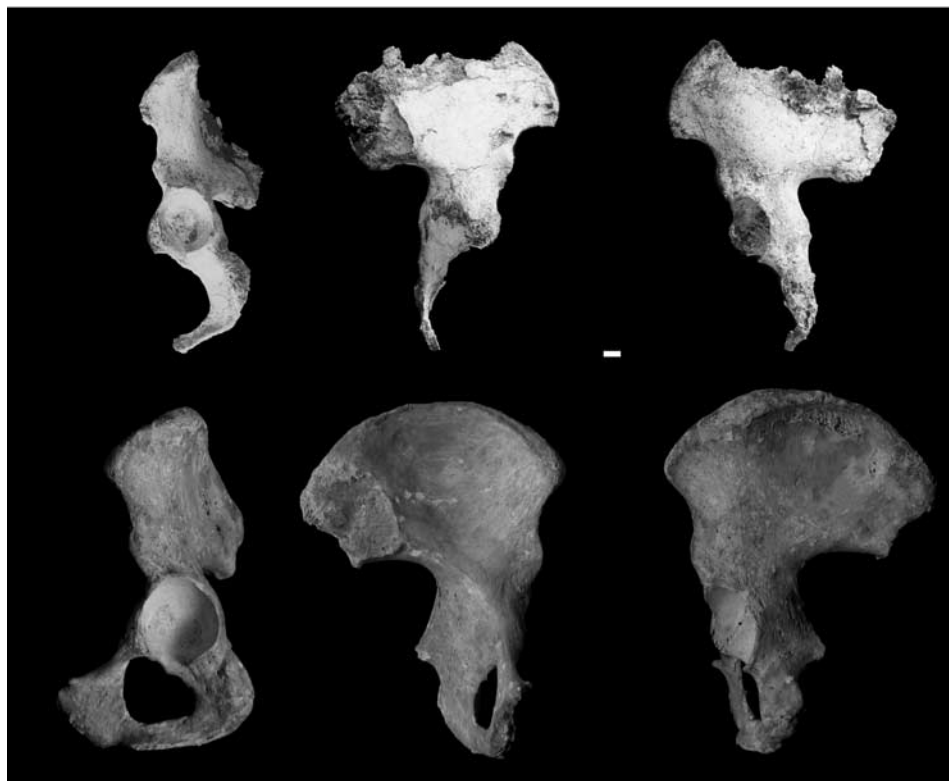


Figure 6 Comparison of the left innominate from LB1 with a modern adult female *H. sapiens*. Lateral (external), and medial and lateral views of maximum iliac breadth. The pubic region of LB1 is not preserved and the iliac crest is incomplete. Scale bar, 1 cm.

an older section of the deposit (74–95 kyr). The estimated maximum length of this radius when complete is approximately 210 mm. Although the arms of LB1 have not been recovered, the dimensions of this radius are compatible with a hominin of LB1 proportions.

Although there is considerable interspecific variation, stature has been shown to have phylogenetic and adaptive significance among hominins²³. Broadly speaking, *Australopithecus* and the earliest members of the genus *Homo* are shorter than *H. erectus* and more recent hominins⁸. The maximum femur length of LB1 (280 mm) is just below the smallest recorded for *A. afarensis* (AL-288-1, 281 mm²⁴) and equal to the smallest estimate for the OH 62 *H. habilis* femur (280–404 mm)²¹. Applying stature estimation formulae developed from human pygmies²⁵ gives a stature estimate of 106 cm for LB1 (Supplementary Information). This is likely to be an overestimation owing to LB1's relatively small cranial height.

A stature estimate for LB1 of 106 cm gives a body mass of 16 to 28.7 kg, and a femur cross-sectional area of 525 mm² gives a mass of 36 kg (Supplementary Information). The brain mass for LB1, calculated from its volume²⁶, is 433.2 g; this gives an encephalization quotient (EQ)²⁷ range of 2.5–4.6, which compares with 5.8–8.1 for *H. sapiens*, 3.3–4.4 for *H. erectus/ergaster* and 3.6–4.3 for *H. habilis*, and overlaps with the australopithecine range of variation^{28,29}. If LB1 shared the lean and relatively narrow body shape typical of Old World tropical modern humans then the smallest body weight estimate, based on Jamaican school children data¹⁹, is probably most appropriate. This would support the higher EQ estimate and place LB1 within the *Homo* range of variation. Although neurological organization is at least as important as EQ in determining behavioural complexity, these data are consistent with *H. floresiensis* being the Pleistocene toolmaker at Liang Bua.

Origins and evolution

The LB1 skeleton was recovered from Flores, an island of 14,000 km² east of the Wallace Line, in Indonesia. It combines extremely small stature and an endocranial volume in the early australopithecine range, with a unique mosaic of primitive and derived traits in the cranium, mandible and postcranial skeleton. Both its geographic location and comparatively recent date suggest models that differ to

those for more expected geological contexts, such as Pliocene eastern Africa. Among modern humans, populations of extremely small average stature were historically found in predominantly rainforest habitat in the equatorial zone of Africa, Asia and Melanesia^{30,31}. Explanations for the small body size of these people generally focus on the thermoregulatory advantages for life in a hot and humid forest, either through evaporative cooling³² or reduced rates of internal heat production³⁰. For African pygmies, smaller body size is the result of reduced levels of insulin-like growth factor 1 (IGF-1) throughout the growth period³³, or reduced receptivity to IGF-1 (ref. 34). Although adult stature is reduced, cranio-facial proportions remain within the range of adjacent larger-bodied populations, as does brain size^{35,36}. The combination of small stature and brain size in LB1 is not consistent with IGF-related postnatal growth retardation. Similarly, neither pituitary dwarfism, nor primordial microcephalic dwarfism (PMD) in modern humans replicates the skeletal features present in LB1 (refs 37–40).

Other mechanisms must have been responsible for the small body size of these hominins, with insular dwarfing being the strongest candidate. Although small body size was an attribute of Pliocene australopithecines, the facial and dental characteristics of LB1 link it with larger-bodied Pleistocene *Homo*. In this instance, body size is not a direct expression of phylogeny. The location of these small hominins on Flores makes it far more likely that they are the end product of a long period of evolution on a comparatively small island, where environmental conditions placed small body size at a selective advantage. Insular dwarfing, in response to the specific ecological conditions that are found on some small islands, is well documented for animals larger than a rabbit^{41,42}. Explanations of the island rule have primarily focused on resource availability, reduced levels of interspecific competition within relatively impoverished faunal communities and absence of predators. It has been argued that, in the absence of agriculture, tropical rainforests offer a very limited supply of calories for hominins⁴³. Under these conditions selection should favour the reduced energy requirements of smaller individuals. Although the details of the Pleistocene palaeoenvironment on Flores are still being documented, it is clear that until the arrival of Mesolithic humans the faunal suit was relatively impoverished, and the only large predators were the Komodo dragon and another larger varanid. Dwarfing in LB1 may have been the end product of selection for small body size in a low calorific environment, either after isolation on Flores, or another insular environment in southeastern Asia.

Anatomical and physiological changes associated with insular dwarfing can be extensive, with dramatic modification of sensory systems and brain size⁴⁴, and certainly exceed what might be predicted by the allometric effects of body size reduction alone. Evidence of insular dwarfing in extinct lineages, or the evolution of island endemic forms, is most often provided by the fossil record. Whereas there is archaeological evidence of hominins being on Flores by approximately 840 kyr⁴⁵, there is no associated hominin skeletal material, and the currently limited evidence from Liang Bua is restricted to the Late Pleistocene. The first hominin immigrants may have had a similar body size to *H. erectus* and early *Homo*^{21,46}, with subsequent dwarfing; or, an unknown small-bodied and small-brained hominin may have arrived on Flores from the Sunda Shelf.

Discussion

When considered as a whole, the cranial and postcranial skeleton of LB1 combines a mosaic of primitive, unique and derived features not recorded for any other hominin. Although LB1 has the small endocranial volume and stature evident in early australopithecines, it does not have the great postcanine tooth size, deep and prognathic facial skeleton, and masticatory adaptations common to members of this genus^{2,47}. Instead, the facial and dental proportions, postcranial anatomy consistent with human-like obligate bipedalism⁴⁸,



Figure 7 Anterior and posterior views of the LB1 right femur and tibia, with cross-sections of the femur neck and midshaft, and tibia midshaft. The anterior surfaces of the medial and lateral condyles of the femur are not preserved. With the exception of the medial malleolus, the tibia is complete and undistorted. Scale bar, 1 cm.

and a masticatory apparatus most similar in relative size and function to modern humans⁴⁸ all support assignment to the genus *Homo*—as does the inferred phylogenetic history, which includes endemic dwarfing of *H. erectus*. For these reasons, we argue that LB1 is best placed in this genus and have named it accordingly.

On a related point, the survival of *H. floresiensis* into the Late Pleistocene shows that the genus *Homo* is morphologically more varied and flexible in its adaptive responses than is generally recognized. It is possible that the evolutionary history of *H. floresiensis* is unique, but we consider it more likely that, following the dispersal of *Homo* out of Africa, there arose much greater variation in the morphological attributes of this genus than has hitherto been documented. We anticipate further discoveries of highly endemic, hominin species in locations similarly affected by long-term genetic isolation, including other Wallacean islands. □

Received 3 March; accepted 8 September 2004; doi:10.1038/nature02999.

- Morwood, M. J. *et al.* Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* doi:10.1038/nature02956 **431**, 1087–1091 (2004).
- Wood, B. A. *Koobi Fora Research Project, Vol. 4: Hominid Cranial Remains* (Clarendon, Oxford, 1991).
- Vekua, A. K. *et al.* A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
- Spoor, C. F. Basicranial architecture and relative brain size of STS 5 (*Australopithecus africanus*) and other Plio-Pleistocene hominids. *S. Afr. J. Sci.* **93**, 182–186 (1997).
- Lieberman, D., Ross, C. F. & Ravosa, M. J. The primate cranial base: ontogeny, function, and integration. *Yearb. Phys. Anthropol.* **43**, 117–169 (2000).
- Baba, H. *et al.* *Homo erectus* calvarium from the Pleistocene of Java. *Science* **299**, 1384–1388 (2003).
- Tobias, P. V. *The Skulls, Endocrasts and Teeth of Homo habilis* (Cambridge Univ. Press, Cambridge, 1991).
- McHenry, H. M. & Coffing, K. E. *Australopithecus to Homo: Transformations of body and mind. Annu. Rev. Anthropol.* **29**, 125–166 (2000).
- Brown, P. Vault thickness in Asian *Homo erectus* and modern *Homo sapiens*. *Courier Forschungs-Institut Senckenberg* **171**, 33–46 (1994).
- Bräuer, G. & Mbua, E. *Homo erectus* features used in cladistics and their variability in Asian and African hominids. *J. Hum. Evol.* **22**, 79–108 (1992).
- Santa Luca, A. P. *The Ngandong Fossil Hominids* (Department of Anthropology Yale Univ., New Haven, 1980).
- Weidenreich, F. The skull of *Sinanthropus pekinensis*: a comparative study of a primitive hominid skull. *Palaentol. Sin.* **D10**, 1–485 (1943).
- Brown, P. & Maeda, T. Post-Pleistocene diachronic change in East Asian facial skeletons: the size, shape and volume of the orbits. *Anthropol. Sci.* **112**, 29–40 (2004).
- Gabunia, L. K. *et al.* Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
- Kaifu, Y. *et al.* Taxonomic affinities and evolutionary history of the Early Pleistocene hominids of Java: dento-gnathic evidence. *Am. J. Phys. Anthropol.* (in the press).
- McHenry, H. M. in *Evolutionary History of the 'Robust' Australopithecines* (ed. Grine, F. E.) 133–148 (Aldine de Gruyter, New York, 1988).
- Wood, B. A. & Yuyterschaut, H. Analysis of the dental morphology of the Plio-Pleistocene hominids. III. Mandibular premolar crowns. *J. Anat.* **154**, 121–156 (1987).
- Wood, B. A., Abbott, S. A. & Yuyterschaut, H. Analysis of the dental morphology of Plio-Pleistocene hominids. IV. Mandibular postcanine root morphology. *J. Anat.* **156**, 107–139 (1988).
- Aiello, A. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, London, 1990).
- Kennedy, G. E. Some aspects of femoral morphology in *Homo erectus*. *J. Hum. Evol.* **12**, 587–616 (1983).
- Hausler, M. & McHenry, H. M. Body proportions of *Homo habilis* reviewed. *J. Hum. Evol.* **46**, 433–465 (2004).
- Stern, J. T. J. & Susman, R. L. The locomotor anatomy of *Australopithecus afarensis*. *Am. J. Phys. Anthropol.* **60**, 279–317 (1983).
- Ruff, C. B. Morphological adaptation to climate in modern and fossil hominids. *Yearb. Phys. Anthropol.* **37**, 65–107 (1994).
- Jungers, W. L. Lucy's limbs: skeletal allometry and locomotion in *Australopithecus afarensis*. *Nature* **297**, 676–678 (1982).
- Jungers, W. L. Lucy's length: stature reconstruction in *Australopithecus afarensis* (A.L.288–1) with implications for other small-bodied hominids. *Am. J. Phys. Anthropol.* **76**, 227–231 (1988).

- Count, E. W. Brain and body weight in man: their antecedents in growth and evolution. *Ann. NY Acad. Sci.* **46**, 993–1101 (1947).
- Martin, R. D. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**, 57–60 (1981).
- Jerison, H. J. *Evolution of the Brain and Intelligence* (Academic, New York, 1973).
- McHenry, H. M. in *The Primate Fossil Record* (ed. Hartwig, C. H.) 401–406 (Cambridge Univ. Press, Cambridge, 2002).
- Cavalli-Sforza, L. L. (ed.) *African Pygmies* (Academic, Orlando, 1986).
- Shea, B. T. & Bailey, R. C. Allometry and adaptation of body proportions and stature in African Pygmies. *Am. J. Phys. Anthropol.* **100**, 311–340 (1996).
- Roberts, D. F. *Climate and Human Variability* (Cummings Publishing Co., Menlo Park, 1978).
- Merimee, T. J., Zapf, J., Hewlett, B. & Cavalli-Sforza, L. L. Insulin-like growth factors in pygmies. *N. Engl. J. Med.* **15**, 906–911 (1987).
- Geffner, M. E., Bersch, N., Bailey, R. C. & Golde, D. W. Insulin-like growth factor I resistance in immortalized T cell lines from African Efe Pygmies. *J. Clin. Endocrinol. Metab.* **80**, 3732–3738 (1995).
- Hiernaux, J. *The People of Africa* (Charles Scribner's Sons, New York, 1974).
- Beals, K. L., Smith, C. L. & Dodd, S. M. Brain size, cranial morphology, climate and time machines. *Current Anthropology* **25**, 301–330 (1984).
- Rimoin, D. L., Merimee, T. J. & McKusick, V. A. Growth-hormone deficiency in man: an isolated, recessively inherited defect. *Science* **152**, 1635–1637 (1966).
- Jaffe, H. L. *Metabolic, Degenerative and Inflammatory Disease of Bones and Joints* (Lea and Febiger, Philadelphia, 1972).
- Seckel, H. P. G. *Bird-Headed Dwarfs* (Karger, Basel, 1960).
- Jeffery, N. & Berkovitz, B. K. B. Morphometric appraisal of the skull of Caroline Crachami, the Sicilian "Dwarf" 1815–1824: A contribution to the study of primordial microcephalic dwarfism. *Am. J. Med. Genet.* **11**, 260–270 (2002).
- Sondaar, P. Y. in *Major Patterns in Vertebrate Evolution* (eds Hecht, M. K., Goody, P. C. & Hecht, B. M.) 671–707 (Plenum, New York, 1977).
- Lomolino, M. V. Body size of mammals on islands: The island rule re-examined. *Am. Nat.* **125**, 310–316 (1985).
- Bailey, R. C. & Headland, T. The tropical rainforest: Is it a productive habitat for human foragers? *Hum. Ecol.* **19**, 261–285 (1991).
- Köhler, M. & Moyà-Solà, S. Reduction of brain and sense organs in the fossil insular bovid *Myotragus*. *Brain Behav. Evol.* **63**, 125–140 (2004).
- Morwood, M. J., O'Sullivan, P. B., Aziz, F. & Raza, A. Fission-track ages of stone tools and fossils on the east Indonesian island of Flores. *Nature* **392**, 173–176 (1998).
- Walker, A. C. & Leakey, R. (eds) *The Nariokotome Homo erectus skeleton* (Harvard Univ. Press, Cambridge, 1993).
- Rak, Y. *The Australopithecine Face* (Academic, New York, 1983).
- Wood, B. A. & Collard, M. The human genus. *Science* **284**, 65–71 (1999).
- Johanson, D. C. & White, T. D. A systematic assessment of early African Hominids. *Science* **202**, 321–330 (1979).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank F. Spoor and L. Aiello for data and discussion. Comments by F. Spoor and D. Lieberman greatly improved aspects of the original manuscript. Conversation with S. Collier, C. Groves, T. White and P. Grave helped clarify some issues. CT scans were produced by CT-Scan KSU, Medical Diagnostic Nusantara, Jakarta. S. Wasisto completed complex section drawings and assisted with the excavation of Sector VII. The 2003 excavations at Liang Bua, undertaken under Indonesian Centre for Archaeology Permit Number 1178/SB/PUS/BD/24.VI/2003, were funded by a Discovery Grant to M.J.M. from the Australian Research Council. UNE Faculty of Arts, and M. Macklin, helped fund the manufacture of stereolithographic models of LB1.

Authors contributions P.B. reconstructed the LB1 cranium and was responsible for researching and writing this article, with M.J.M. T.S. directed many aspects of the Liang Bua excavations, including the recovery of the hominin skeleton. M.J.M. and R.P.S. are Principal Investigators and Institutional Counterparts in the ARC project, as well as Co-Directors of the Liang Bua excavations. E.W.S. and Jatmiko assisted T.S., and had prime responsibility for the work in Sector VII. R.A.D. did all of the initial faunal identifications at Liang Bua, including hominin material, and helped clean and conserve it.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.B. (pbrown3@pobox.une.edu.au).

Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism

Fadel A. Samatey^{1,2,3}, Hideyuki Matsunami^{1,2,3}, Katsumi Imada^{1,2,3}, Shigehiro Nagashima^{1,3}, Tanvir R. Shaikh^{4*}, Dennis R. Thomas⁴, James Z. Chen⁴, David J. DeRosier⁴, Akio Kitao⁵ & Keiichi Namba^{1,2,3}

¹Dynamic NanoMachine Project, ICORP, JST, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan

²Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan

³Protonic NanoMachine Project, ERATO, JST, 3-4 Hikaridai, Seika, Kyoto 619-0237, Japan

⁴W. M. Keck Institute of Cellular Visualization, Rosenstiel Basic Medical Sciences Research Center and Department of Biology, Brandeis University, Waltham, Massachusetts 02454-9110, USA

⁵Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan

* Present address: Wadsworth Center, P.O. Box 509, Albany, New York 12201, USA

The bacterial flagellum is a motile organelle, and the flagellar hook is a short, highly curved tubular structure that connects the flagellar motor to the long filament acting as a helical propeller. The hook is made of about 120 copies of a single protein, FlgE, and its function as a nano-sized universal joint is essential for dynamic and efficient bacterial motility and taxis. It transmits the motor torque to the helical propeller over a wide range of its orientation for swimming and tumbling. Here we report a partial atomic model of the hook obtained by X-ray crystallography of FlgE31, a major proteolytic fragment of FlgE lacking unfolded terminal regions, and by electron cryomicroscopy and three-dimensional helical image reconstruction of the hook. The model reveals the intricate molecular interactions and a plausible switching mechanism for the hook to be flexible in bending but rigid against twisting for its universal joint function.

To propel themselves in their living environments towards favourable conditions and away from unfavourable ones, bacteria have developed highly sophisticated machinery called the flagellum. It is a complex molecular machine made of about 25 different proteins, each in multiple copies from a few to a few tens of thousands^{1–3}. The flagellum can be divided into three parts: the basal body, the hook and the filament^{4,5}. The basal body is a rotary motor^{6,7}, and its complex structure, which is made of about 20 proteins, begins inside the cell, spans the cell envelope including the cytoplasmic membrane, and extends well outside the cell^{1,8}. The filament, a long tubular structure, is a helical assembly of some tens of thousands of copies of a single protein FlgC (flagellin). Its polymorphic supercoiled forms permit it to function as a helical propeller, which switches its helical pitch and handedness depending on the swimming mode^{9,10}. The hook, a short, highly curved tube, is a helical assembly of about 120 copies of a single protein, FlgE (refs 4, 5, 11, 12), also called the hook protein. The hook connects the basal body to the filament. Its flexibility permits it to transmit torque from the motor to the helical propeller when the two are not coaxial⁶. The hook allows the synchronous rotation of several filaments driven by their motors in a bundle formed behind the cell (swimming) as well as the uncoordinated rotation of individual, unbundled filaments in different orientations^{6,9,10} (tumbling). An appropriate length and bending flexibility of the hook seem to be important for its universal joint function¹³. There are also two proteins, hook-associated protein 1 (HAP1) and 3 (HAP3), forming a very short hook-filament junction, which probably acts as a structural adaptor for a smooth transition between the two mechanically distinct structures: the hook is relatively flexible in bending, whereas the filament is much more rigid for its propeller function.

The hook of *Salmonella* flagella and its component protein FlgE have been studied by electron microscopy as well as by biochemical and physicochemical methods. The length of the hook is relatively well regulated to be 55 ± 6 nm (ref. 14), but it becomes much

longer if there is a mutation in FliK or FlhB, two proteins involved in protein export and flagellar assembly. The helical packing of subunits in this abnormally long hook, called a polyhook, is the same as that in the normal hook¹⁵. This helical packing is distorted into a superhelical form, although with an amplitude and pitch smaller than those of the superhelical flagellar filament. The wild-type hook is thus a short segment of this superhelical form rather than a simply bent rod structure. The polyhook also shows polymorphism, transforming into distinct helical forms as well as a straight form depending on the solution condition such as pH, ionic strength and temperature^{12,16–18}. The structures of various straight polyhooks studied by electron microscopy and helical image analysis have shown that the basic architecture of the hook is similar to that of the flagellar filament, which can be described as a tubular fibre made of 11 protofilaments or a helical symmetry of about 11 subunits per two turns of the 1-start helix^{11,19}, in spite of the fact that the hook protein and flagellin have very different amino acid sequences²⁰.

To understand the mechanism of this molecular universal joint, we have solved the structure of a core fragment of FlgE by X-ray crystallography at 1.8 Å resolution and built an atomic model of the hook by docking it into a density map of the hook obtained by electron cryomicroscopy and image analysis. The hook model now shows its complex molecular interactions and a possible switching mechanism to form a highly curved and twisted tubular structure within which individual protofilaments go through rather large and dynamic conformational changes for their rather extensive elongation and shortening as the curved hook rotates rapidly.

Structure of FlgE31

Because full-length FlgE, like FliC (flagellin), forms filaments not crystals, we cloned, overexpressed in *Escherichia coli* and crystallized a fragment named FlgE31, corresponding to residues 71–369 out of 402 and having a molecular mass of 31 kDa. This same trick works

for FlhC (refs 21, 22). This fragment, which lacks both terminal regions that are unfolded in the monomeric form in solution²³, has two compact domains as judged from its calorimetric melting profile²⁴. Indeed, the C α backbone trace of FlgE31 revealed by X-ray crystallography actually shows two domains, D1 and D2, connected by a short stretch of two-strand anti-parallel β -sheet (Fig. 1). The orientation of FlgE31 in Fig. 1 (namely D1 at the bottom and D2 at the top) is that found in a hook whose cell proximal side is down; this orientation will be used throughout the paper.

In the atomic model, the amino-terminal segment from Gly 71 to Ala 144 and the carboxy-terminal segment from Pro 285 to Ser 363 comprise D1, and a central segment from Ala 145 to Lys 284 makes up domain D2. The last six residues of FlgE31 were invisible in the electron density map. Both domains have an oval shape and are made mostly of β -structures. Domain D2 is a flattened, eight-strand β -barrel but with significant irregularity and extra loops. Domain D1 has a rather complex, unusual fold composed of many different folding motifs: a stack of four horizontal β -hairpins one above another, alternating their orientations with crossing angles of about 120° (Asn 79–Leu 115 and Gly 324–Gln 337, on the left side in Fig. 1); a triangular loop (Thr 116–Pro 135, on the right side in front); a four-stranded (Leu 288–Ile 314 and Asn 357–Ser 363) β -sheet (in the upper and lower half, respectively, both on the back side); two consecutive β -turns (Thr 346–Phe 352, behind the triangular loop); and a vertically extended chain (Pro 135–Ala 144, in the centre front of the upper half). This extended chain seems to be a backbone around which the other motifs assemble. A three-dimensional structural similarity search using software DALI²⁵ resulted in no match for domain D1, confirming its unique fold. The longest dimensions of domain D1 and D2 are about 50 and 45 Å, respectively, and these two domains are connected along their long axes with an angle of about 70°.

As predicted from amino acid sequences and expected from far-ultraviolet circular dichroic spectra, the structure of FlgE31 is very

different from that of the F41 fragment of flagellin^{21,26}, which consists of three domains with domain D1, which consists of three α -helices and a β -hairpin, domain D2, which is formed from many β -hairpins, and domain D3, which is made of a tight β -barrel. It is curious that these two molecules with completely different structures both form the tubular structures with basically the same architecture and helical symmetry.

Partial atomic model of the hook

To build a model of the hook, we then obtained a three-dimensional image reconstruction of a straight hook by using electron cryo-microscopy and image analysis, and docked the crystal structure of FlgE31 in the density map, the details of which will be described elsewhere (T.R.S., D.R.T., J.Z.C., F.A.S., H.M., K.I., K.N., D.J.D., manuscript in preparation). The density map of the hook showed three domains: the outermost domain on the surface of the hook at a radius of 7.5 nm, the middle domain that lies between radii of 5 and 6 nm, and inner core domain that forms a tube with a wall about 1 nm thick and a 3-nm axial lumen. The inner core domain, which is rod-shaped, about 1 nm in diameter and about 2.5 nm long, is most likely to be an α -helical coiled coil made of both terminal chains of FlgE, just like the one seen in the atomic model of flagellin in the filament²⁶. The two-domain structure of FlgE31 was therefore docked into the middle and outer domains of the hook density map.

The initial docking was performed manually by fitting each domain separately, with domain D2 into the outer density feature and D1 into the middle one, on the basis of the reasonable assumption that the terminal chains of FlgE are located in the inner core of the hook in a similar way to those of flagellin in the filament²⁶. The two domains were then connected, and the hook model was refined with a real-space structure refinement program²⁷. The atomic model of the hook in a C α backbone ribbon diagram is superimposed on the density map in Fig. 2a, b, showing that the model fits nicely in the density. The correctness of this hook model

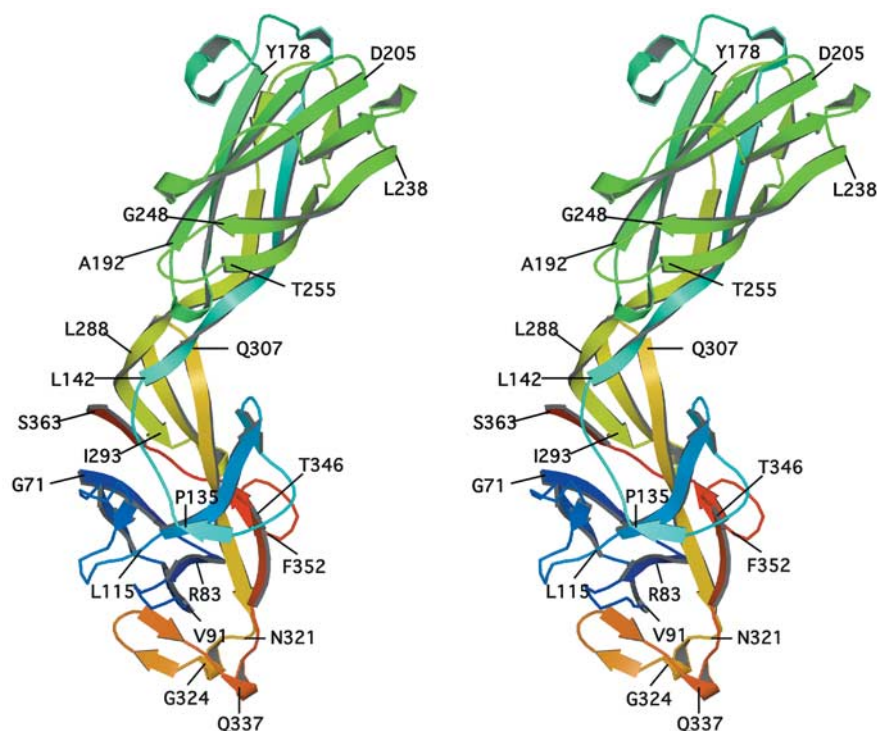


Figure 1 Stereo view of the C α backbone trace of FlgE31. The chain is colour-coded from blue to red, going through the rainbow colours, from the N terminus to the C terminus. The model is oriented with domain D1 at the bottom and D2 at the top. Domain D1 shows a

stack of four β -hairpins on the left and a triangular loop on the front right. All figures were prepared with MOLSCRIPT⁴⁵ and RASTER3D⁴⁶.

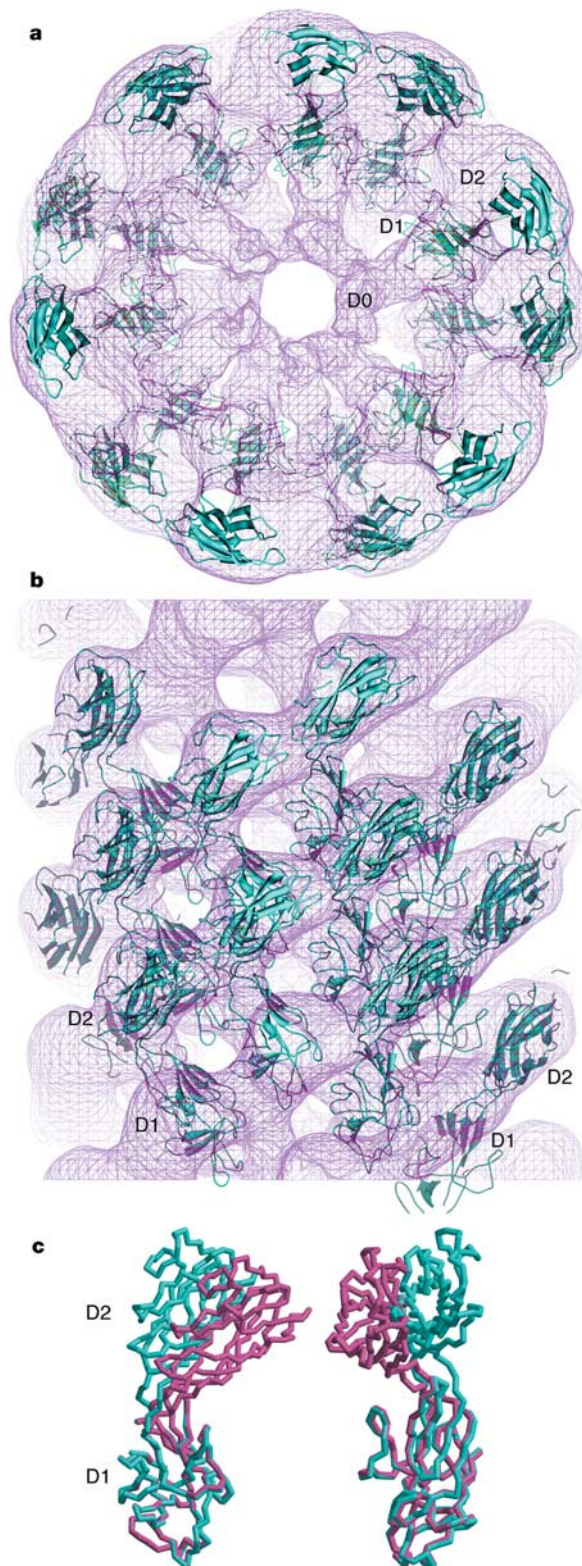


Figure 2 Docking of the atomic model of FlgE31 into the outer two domains of the hook. **a**, End-on view. **b**, Side view. In the side view, slightly more than half of the hook density on the back is trimmed off for clarity. **c**, Change in the relative domain arrangement involved in the docking and refinement. Two different views are presented. The models are colour-coded as follows: purple, crystal structure; cyan, refined model.

will be discussed in more detail elsewhere (T.R.S., D.R.T., J.Z.C., F.A.S., H.M., K.I., K.N., D.J.D, manuscript in preparation). The relative orientations of the major axes of domains D1 and D2 had to be changed for the docking and refinement as shown in Fig. 2c. The change in angle between the major axes is small but the two domains had to be twisted with respect to each other by about 10° , indicating that the hinge at the D1–D2 connection is relatively flexible. This flexibility in the subunit conformation seems to have an important function in the mechanical property of the hook, as discussed below.

Interactions between subunits

It has been indicated even in density maps of relatively low resolution^{11,19} that the packing interactions of D2 domains on the hook surface are strong along the 6-start helix, which is a right-handed helical line tilted about 50° to the hook axis. The atomic model as well as the new image reconstruction at higher resolution now shows extensive contacts between D2 domains along the 6-start helix (Figs 2 and 3). D2 domain contacts are found neither along the 11-start helix, which is a nearly axial line, nor the 5-start helix, which is a left-handed helical line at a tilt angle of about 60° (Fig. 4a).

The intersubunit packing interactions between D1 domains on the inside of the straight hook are not very extensive in any of the three helical directions described above, but show more or less equally weak contacts between side chains (Fig. 4a). The top and bottom portions of the stacked β -hairpins have interactions along the 5-start and 6-start, respectively, with the C-terminal chains on the lower and upper sides of domain D1, respectively. The top to bottom interaction along the 11-start is made between the C-terminal chains. The residues involved in these interactions are summarized in Supplementary Table 1.

Along the 11-start helices there are intimate intersubunit interactions between domain D2 of the lower subunit and domain D1 of the upper subunit through the triangular loop of domain D1 (Fig. 4b), where three side-chain contacts are observed (Supplementary Table 1). This is in contrast to the 11-start intersubunit interactions of the flagellar filament. The 11-start direction defines the protofilaments of the flagellar filament, which are the cooperative units that switch between the L and R states to produce the superhelical form of the filament. In the flagellar protofilament, the D1–D1 interactions are much tighter, involving intermolecular β -structure, and its rigidity, which resists axial extension or compression, is important in maintaining the two distinct intersubunit repeat distances of the L and R conformations rather strictly at 52.7 and 51.9 Å, respectively. The presence of both L and R protofilaments in the filament gives rise to its superhelical form; the rigidity of the form is essential to its function as a helical propeller. In contrast, the axial D1–D1 interaction in the hook is only a weak contact involving a few side chains, and its 11-start protofilament structure is mainly held by the D1–D2 interactions with these two domains arranged radially. This is important for its role as a universal joint mechanism as described below.

All the interactions found in the hook model made of domains D1 and D2 are either polar–polar or polar–charge interactions, explaining why FlgE31 cannot polymerize into a stable hook-like fibre structure in aqueous solution. Close interactions between terminal chains in domain D0 in the inner core of the hook are responsible for the stable hook structure formation in a similar way to those of the inner core of the filament²⁶.

Model of a curved hook

The wild-type hook is highly curved as observed by electron microscopy⁴. Hook protofilaments on the inside of the curve must therefore have shorter repeat distances than those on the outside of the curve, and the difference is much greater than that of the flagellar protofilaments²⁸. Although acting as a universal joint, the curvature and curve–linear axis of the hook seem stationary

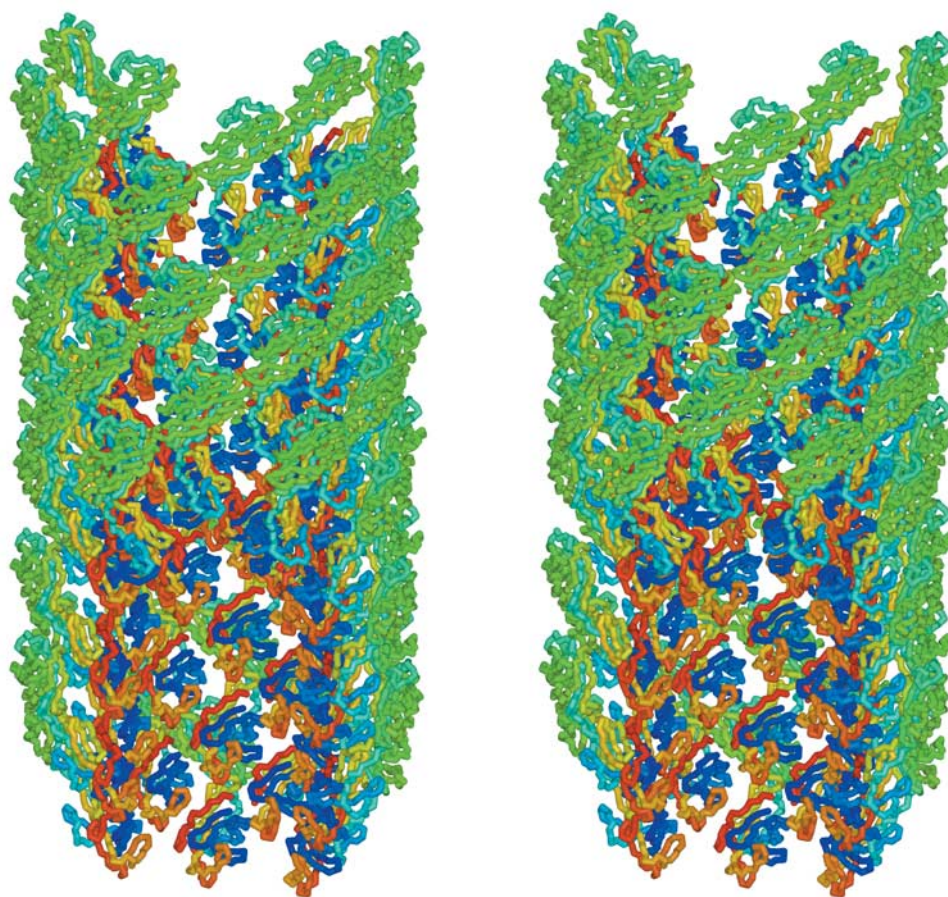


Figure 3 Stereo view of the atomic model of the D1–D2 part of the straight hook. The C α backbone of each subunit is colour-coded as in Fig. 1. Three of the 11-stranded

protofilaments are removed in the front part for the bottom half and in the back for the top half, for clearer views of domain interactions.

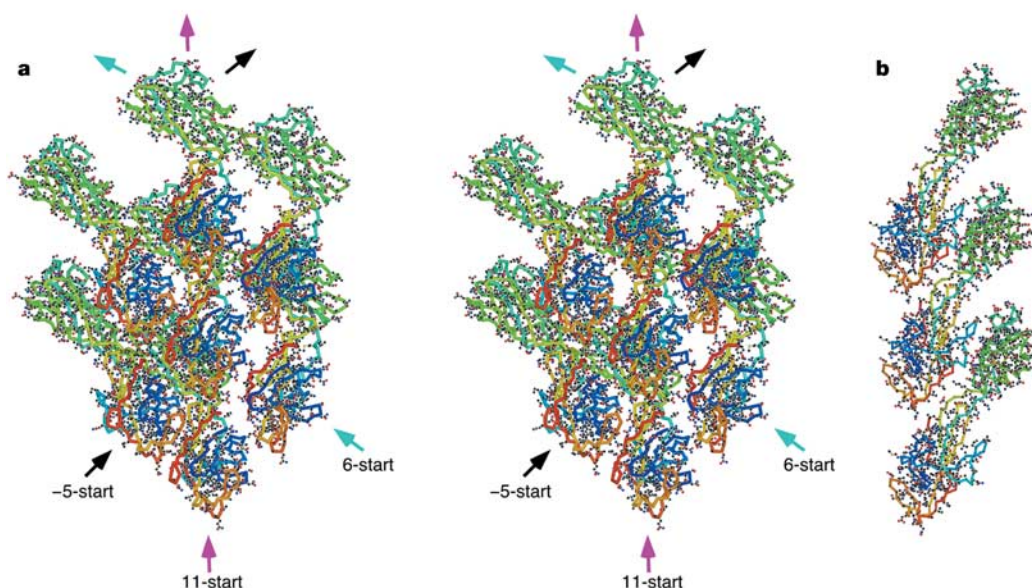


Figure 4 Magnified views of intermolecular interactions along various helical lines of the straight hook. **a**, Seven subunits viewed from the inside (stereo view). **b**, Three subunits along the protofilament (11-start). Atomic models are represented with the C α backbone

in stick form and side chains in ball-and-stick form, and colour-coded as in Fig. 1. Arrows in three different colours in **a** indicate the directions of three representative helical lines: the bottom three for D1 domains and the top three for D2 domains.

within the plane containing the axes of the motor and filament; yet, as it rotates, the inside of the bend is successively occupied by different protofilaments. Thus during this kind of rolling rotation, which is like the rotation of smoke in a smoke ring, hook protofilaments must have continuously varying repeat distances and must be going through dynamic conformational changes: an axial extension and a compression with every revolution (Supplementary Videos 1 and 2), occurring about 300 times a second^{29,30}. To see the amounts of axial extension and compression and what types of structural change would be responsible for them, we built a preliminary model of a curved hook based on the atomic model of the straight hook. We continuously deformed the helical lattice of the straight hook so that the hook axis conforms a right-handed helical line having a pitch of 950 Å and a diameter of 350 Å as observed previously¹⁸.

The model of the curved hook is shown in Fig. 5a, with a schematic diagram of the basal body at its base. The axial distances between FlgE subunits are made shorter on the inner side and longer on the outer side than those of the straight hook. A short segment of the curved hook is magnified and two 11-start helical arrays, one on the inside and the other on the outside, are shown in Fig. 5b. The near-axial repeat distances between D2 domains on the inside and outside of the curved hook are 35 and 59 Å respectively, whereas those between D1 domains are 39 and 54 Å. In the straight hook these distances are all 46 Å (refs 11, 19); thus, the interdomain distances along the protofilament have to be compressed or extended by about 6–8 Å for the D1 array and by 11–13 Å for the D2 array. The predicted changes in the domain packing are shown magnified in Fig. 5c, d. This is in contrast to the difference in the repeat distance of only 0.8 Å in the protofilament of the filament as mentioned above³¹. How could this large conformational change of the hook protofilament be possible?

The main axial intermolecular interactions that hold the protofilament structure are between domain D1 of the upper subunit and D2 of the lower subunit through the triangular loop of domain D1, which looks like the ampersand character ‘&’ (Figs 4b and 5b). If

domains D1 and D2 do not change conformation during rotation of the hook, the only plausible mechanisms for this rather large change in the protofilament repeat distance are either a large conformational change of the triangular loop, which is relatively isolated and independent from the hydrophobic core of domain D1, or a large slip at the interface between the triangular loop and the inner face of domain D2.

We therefore performed a molecular dynamic simulation of extension and compression of the protofilament in a similar way to the simulation that we performed for the flagellar protofilament²¹. We used the atomic model of the isolated hook protofilament made of either two or three subunits, but this time not in a vacuum but surrounded by many water molecules because the model had rather large gaps between domains of interacting subunits. The simulation was done by shifting the reference coordinates of domain D2 of the subunit at the top and domain D1 of the subunit at the bottom in opposite directions and applying positional restraints to the main-chain atoms of these domains to compress by 5 Å or extend by 15 Å (Fig. 6). The conformation of the triangular loop shows relatively small changes after axial compression or extension; instead, the bonding interactions between amino acid residues of the triangular loop and the inner face of D2 show multiple steps of changes in bonding partners (Fig. 6b), realizing a large slippage at this D1–D2 interface, just as shown in Fig. 5b. Certain flexibility in the relative domain orientation between D1 and D2, which was indicated in the docking process of the crystal structure of FlgE31 into the hook density map, seems to have a function in the changes in bonding partners. The bending flexibility of the hook, which is essential for its universal joint function and has actually been observed in electron micrographs of the isolated filament–hook–basal-body complex⁴, is probably due to this stepwise axial sliding along with flexibility in relative domain orientation.

Polymorphic supercoiling of the hook

A well-defined switching of the protofilament repeat by 0.8 Å and a

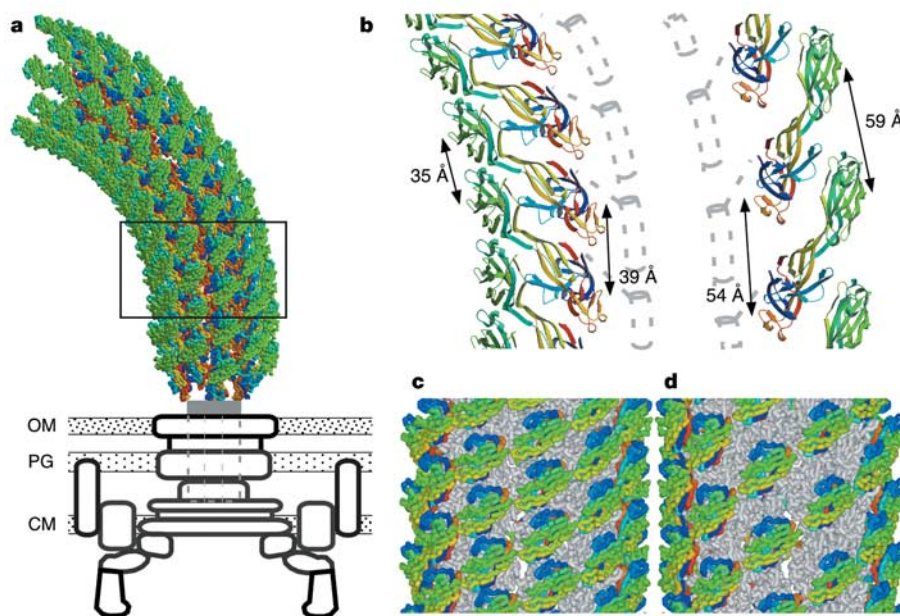


Figure 5 Atomic model of the supercoiled hook. **a**, Atomic model of the coiled hook with a schematic diagram of the basal body spanning the outer membrane (OM) and the cytoplasmic membrane (CM) as well as the peptidoglycan layer (PG). This coiled hook model is part of a supercoiled polyhook with a helical pitch of 950 Å and a diameter of 350 Å. **b**, Magnified image of the coiled hook with the innermost and outermost

protofilaments on the left and right, respectively. The inner core domains formed from both terminal chains and the central channel are represented by dotted grey lines. **c**, **d**, Intermolecular packing arrangements of D2 domains on the inner side (**c**) and on the outer side (**d**) of the coiled hook surface. Only domain D2 is colour-coded as in Fig. 1; domain D1 is coloured light grey.

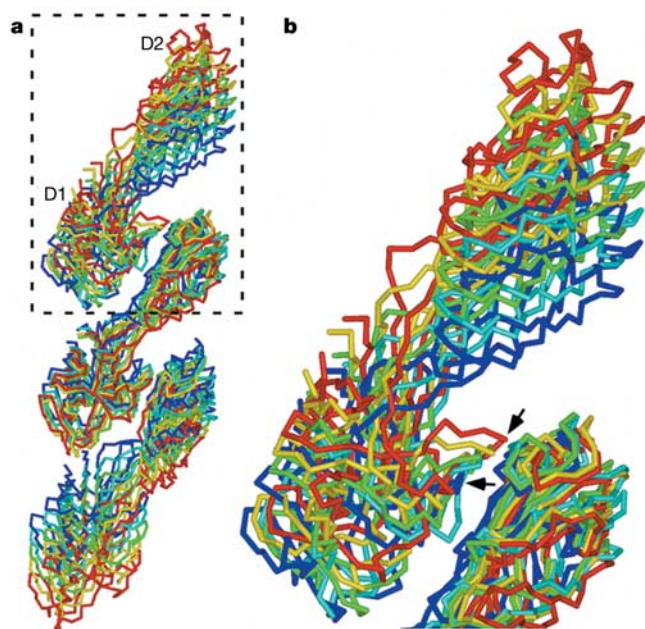


Figure 6 Simulated extension and compression of hook protofilament. Three subunits were used, and domain D2 at the top subunit was moved upwards whereas domain D1 of the bottom subunit was moved downwards. Protofilament models at five different stages at 5 Å intervals are superimposed with different colours: dark blue, light blue, green, yellow and red, from the most compressed to the most extended state. **a**, The whole three subunits. **b**, Magnified view of upper half. D2 domains at the top and D1 domains at the bottom are at equal intervals (2.5 Å) along the vertical axis. It should be noted that the top corner of the triangular loop of domain D1 of the top subunit (indicated by arrows for the most compressed and the most extended, respectively) shows stepwise movements (a jump from dark blue to light blue, very small movements from light blue to green and to yellow, and another jump to red), whereas domain D2 of the middle subunit stays more or less in the same position. This indicates that the triangular loop of domain D1 and the surface of domain D2 can have distinct side-chain bonding partners depending on the state of extension or compression.

relative sliding of neighbouring protofilaments by 2.6 Å are tightly coupled to produce limited sets of twist and curvature for supercoiled flagellar filaments³¹. This is the basis for the polymorphic supercoiling of the filament, in which the cooperatively switching protofilament is important. Kato *et al.*¹⁸ also showed that the polyhook also goes through polymorphic transformations of its supercoiled form in response to changes in the salt concentration, pH and temperature of the solution. Their twist–curvature diagram obtained from the pitch and diameter of helical polyhooks indicated that a two-state model similar to that of the filament might also be valid for the hook, but the inclination angle of the cooperatively switching protofilament to the hook axis in the straight hook is about 50°, much larger than for either of the two types of straight flagellar filament. Assuming a diameter of 200 Å for the hook, which is slightly larger than observed, they suggested that the protofilament must lie along the left-handed 16-start family, but no physical contacts between domains are found along this helical array in the hook model. As shown in the hook model in Figs 2 and 3a, the left-handed 5-start helical array of D1 domains is about 50° to the hook axis; this could therefore be the protofilament for supercoiling. The problem is that the normal polyhook is a right-handed helix, and to make the 5-start helical family have a right-handed inclination the neighbouring domain arrays have to slide to each other over a distance corresponding to about one subunit, which is not a physically plausible mechanism.

An alternative and more plausible mechanism is suggested from the close packing arrangement of D2 domains on the inner side of

the curved hook model. If the curvature of the hook is produced by close packing interactions between D2 domains on the inner side while the overall axial repeat of subunits is kept constant by the axial packing interactions of D0 domains in the inner core, which is missing from the present model, a few sets of distinct D2–D2 interactions can define a limited number of virtual protofilament directions to produce distinct supercoils easily, in both left-handed and right-handed forms. The curvature and twist of each supercoil would depend on the direction of the D2–D2 interaction, producing a well-defined distribution of points in the twist–curvature diagram, although it might not necessarily be on a sinusoidal curve as has been observed for the supercoiled filaments. The currently available data do not seem to be sufficient to determine which model is correct. More precise measurements of the pitch and diameter of the polyhooks of various helical forms are necessary to answer this question.

In any case, the structural and functional differences between the hook and the filament are clearly demonstrated here, illustrating the importance and the function of the hook-associated proteins HAP1 and HAP3, which form the hook–filament junction. To join the hook to the filament, HAP1 and HAP3 presumably operate as a structural and mechanical adaptor. Indeed, mutations in HAP3 cause the rotating filament to lose its superhelicity under stress³². Although it is not obvious why two proteins are necessary, the structures of HAP1 and HAP3 are expected to have some similarity to those of the hook protein and flagellin, respectively. These two structures have been solved by X-ray crystallography and will be described elsewhere (K.I., H.M., F.A.S., S.N. and K.N., manuscript in preparation).

Methods

Preparation and crystallization of FlgE31

The preparation and crystallization of FlgE31 from *Salmonella typhimurium* will be described elsewhere³³. In brief, with the use of the hanging drop vapour diffusion method, crystals were obtained by equilibrating a protein solution containing 1.5 mg ml^{−1} FlgE31, 6% PEG-2000, 1.5 mM cuprous acetate, 25 mM sodium cacodylate, with a reservoir solution consisting of 12% poly(ethylene glycol) (2000 Da), 3 mM cuprous acetate, 50 mM sodium cacodylate, pH 4.5. Crystals appeared after 10 days and the biggest crystal grew to about 0.8 mm × 0.2 mm × 0.3 mm.

Data collection and structure analysis

FlgE31 crystals were frozen in liquid propane equilibrated with liquid nitrogen. All the data sets were collected at temperatures of about 100 K. The native data set was collected at beamline ID29 at the European Synchrotron Radiation Facility, Grenoble. The space group of the crystal is *P*2₁2₁2, with cell dimensions *a* = 128.7 Å, *b* = 49.0 Å, *c* = 96.9 Å. The solvent content is 48% and the crystal contains two molecules per asymmetric unit. A set of multi-wavelength anomalous diffraction (MAD) data, which produced a high-quality electron density map at 2.3 Å resolution, were collected at beamline BL41XU at the 8-GeV Super Photon ring (SPring-8) in Harima. The data were reduced by DENZO and SCALEPACK³⁴, or MOSFLM³⁵ and SCALA³⁶. The phases at 2.3 Å were obtained from the set of MAD data and SOLVE³⁷. The phases were extended by density modification to 1.8 Å with RESOLVE^{37,38}. Partial tracing of the chain was performed with ARP/WARP³⁹ and completed with O⁴⁰. The model was refined at 1.8 Å resolution including 763 water molecules (Supplementary Table 2) with REFMAC5 (refs 36, 41). In the final model of FlgE31, the six last residues in the C-terminal chain could not be traced because of the poor quality of the density map at this end.

Electron cryomicroscopy and helical image analysis

Polyhooks were isolated as described previously⁴². Grids were prepared for electron cryomicroscopy with the use of protocols for flagellar filaments⁴³ except that grids were prepared at 4 °C, which straightened about 50% of the normally curly hooks. Electron microscopy was performed at 200 keV with a field emission gun at about × 65,000 magnification and a range of underfocus from 1.3 to 2.7 μm and a dose of 10 electrons Å^{−2}. We cut digitized images (pixel size 3.2 Å) of frozen-hydrated polyhooks into 420 total polyhook segments of about 800 hook subunits. The segments that failed to reveal at least three layer-lines in their Fourier transforms were subsequently rejected. Layer-line data sets from the remaining 354 images were aligned and merged by using cross-correlation methods. Further details of the helical image analysis, three-dimensional image reconstruction, and docking of the atomic model into the density will be fully described elsewhere (T.R.S., D.R.T., J.Z.C., F.A.S., H.M., K.I., K.N., D.J.D., manuscript in preparation; details are available from the authors on request). Fourier data out to 9 Å resolution have been included in the three-dimensional image reconstruction although the true resolution may not be as good as that.

Simulated extension and compression of the protofilament

Molecular dynamics simulation was performed by using the module SANDER in the molecular simulation program package AMBER7 (ref. 44) with the parm99 force field. As the isolated hook protofilament model, the atomic model consisting of two or three FlgE31 subunits was initially placed in a rectangular box. The gaps were filled with water molecules and counterions. To allow the large extension of inter-subunit distances, the margin from the subunit to the boundary was initially set to be at least 25 Å along the z-axis. The initial box size, the number of water molecules, sodium ions and total atoms were 89.1 Å × 81.3 Å × 191.6 Å, 36,567, 18 and 118,149 in the two-subunit simulation and 92.6 Å × 89.2 Å × 237.9 Å, 52,165, 27 and 169,167 in the three-subunit simulation, respectively. Periodic boundary conditions were used and non-bonded interactions were calculated by the particle-mesh Ewald method. After equilibration in an isothermal–isobaric ensemble at 300 K and 1 atm, the reference coordinates of domain D2 of the subunit at the top and domain D1 of the subunit at the bottom were shifted in opposite directions along the protofilament axis, and positional restraints were applied to the main-chain atoms so as to displace these domains to the position of the reference coordinates. The stepwise shift of the reference coordinates was made by 1 Å, up to 15 Å in extension and up to 5 Å in compression. In each step, 10-ps molecular dynamics simulation was performed and the final coordinate was used for the analysis.

Received 1 July; accepted 8 September 2004; doi:10.1038/nature02997.

- Macnab, R. M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).
- Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).
- Kojima, S. & Blair, D. The bacterial flagellar motor: structure and function of a complex molecular machine. *Int. Rev. Cytol.* **233**, 93–134 (2004).
- DePamphilis, M. L. & Adler, J. Purification of intact flagella from *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **105**, 376–383 (1971).
- DePamphilis, M. L. & Adler, J. Fine structure and isolation of the hook–basal body complex of flagella from *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **105**, 384–395 (1971).
- Berg, H. C. & Anderson, R. A. Bacteria swim by rotating their flagellar filaments. *Nature* **245**, 380–382 (1973).
- Silverman, M. & Simon, M. Flagellar rotation and the mechanism of bacterial motility. *Nature* **249**, 73–74 (1974).
- Namba, K. & Vonderviszt, F. Molecular architecture of bacterial flagellum. *Q. Rev. Biophys.* **30**, 1–65 (1997).
- Macnab, R. M. & Ornston, M. K. Normal-to-curly flagellar transitions and their role in bacterial tumbling. Stabilization of an alternative quaternary structure by mechanical force. *J. Mol. Biol.* **112**, 1–30 (1977).
- Turner, L., Ryu, W. S. & Berg, H. C. Real-time imaging of fluorescent flagellar filaments. *J. Bacteriol.* **182**, 2793–2801 (2000).
- Wagenknecht, T., DeRosier, D. J., Aizawa, S.-I. & Macnab, R. M. Flagellar hook structures of *Caulobacter* and *Salmonella* and their relationship to filament structure. *J. Mol. Biol.* **162**, 69–87 (1982).
- Kagawa, H., Aizawa, S. I. & Asakura, S. Transformations in isolated polyhooks. *J. Mol. Biol.* **129**, 333–336 (1979).
- Williams, A. W. *et al.* Mutation in fliK and flhB affecting flagellar hook and filament assembly in *Salmonella typhimurium*. *J. Bacteriol.* **178**, 2960–2970 (1996).
- Hirano, T., Yamaguchi, S., Oosawa, K. & Aizawa, S.-I. Roles of FliK and FlhB in the determination of flagellar hook length in *Salmonella typhimurium*. *J. Bacteriol.* **176**, 5439–5449 (1994).
- Wagenknecht, T., DeRosier, D. J., Shapiro, L. & Weissborn, A. Three-dimensional reconstruction of the flagellar hook from *Caulobacter crescentus*. *J. Mol. Biol.* **151**, 439–465 (1981).
- Kutsukake, K., Suzuki, T., Yamaguchi, S. & Iino, T. Role of gene flaFV on flagellar hook formation in *Salmonella typhimurium*. *J. Bacteriol.* **140**, 267–275 (1979).
- Aizawa, S. & Maeda, Y. A new method for determination of parity in optical diffraction patterns from the structures with helical symmetry. *J. Mol. Biol.* **137**, 437–442 (1980).
- Kato, S., Okamoto, M. & Asakura, S. Polymorphic transition of the flagellar polyhook from *Escherichia coli* and *Salmonella typhimurium*. *J. Mol. Biol.* **173**, 463–476 (1984).
- Morgan, D. G., Macnab, R. M., Francis, N. R. & DeRosier, D. J. Domain organization of the subunit of the *Salmonella typhimurium* flagellar hook. *J. Mol. Biol.* **229**, 79–84 (1993).
- Homma, M., DeRosier, D. J. & Macnab, R. M. Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. *J. Mol. Biol.* **213**, 819–832 (1990).
- Samatey, F. A. *et al.* Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* **410**, 331–337 (2001).
- Samatey, F. A., Imada, K., Vonderviszt, F., Shirakihara, Y. & Namba, K. Crystallization of the F41

- fragment of flagellin and data collection from extremely thin crystals. *J. Struct. Biol.* **132**, 106–111 (2000).
- Vonderviszt, F., Ishima, R., Akasaka, K. & Aizawa, S. Terminal disorder: a common structural feature of the axial proteins of bacterial flagellum? *J. Mol. Biol.* **226**, 575–579 (1992).
- Vonderviszt, F., Zavodszky, P., Ishimura, M., Uedaira, H. & Namba, K. Structural organization and assembly of flagellar hook protein from *Salmonella typhimurium*. *J. Mol. Biol.* **251**, 520–532 (1995).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650 (2003).
- Chen, J. Z., Fürst, J., Chapman, M. S. & Grigorieff, N. Low resolution structure refinement in electron microscopy. *J. Struct. Biol.* **144**, 144–151 (2003).
- Hasegawa, K., Yamashita, I. & Namba, K. Quasi- and nonequivalence in the structure of bacterial flagellar filament. *Biophys. J.* **74**, 569–575 (1998).
- Lowe, G., Meister, M. & Berg, H. C. Rapid rotation of flagellar bundles in swimming bacteria. *Nature* **325**, 637–640 (1987).
- Kudo, S., Magariyama, Y. & Aizawa, S. Abrupt changes in flagellar rotation observed by laser dark-field microscopy. *Nature* **346**, 677–680 (1990).
- Yamashita, I. *et al.* Structure and switching of bacterial flagellar filaments studied by X-ray fiber diffraction. *Nature Struct. Biol.* **5**, 125–132 (1998).
- Fahrner, K. A., Block, S. M., Krishnaswamy, S., Parkinson, J. S. & Berg, H. C. A mutant hook-associated protein (HAP3) facilitates torsionally induced transformations of the flagellar filament of *Escherichia coli*. *J. Mol. Biol.* **238**, 173–186 (1994).
- Samatey, F. A., Matsunami, H., Imada, K., Nagashima, S. & Namba, K. Crystallization of a core fragment of the hook protein FlgE. *Acta Crystallogr. D* (in the press).
- Otwinowski, Z. & Minor, W. *Processing of X-ray Diffraction Data Collected in Oscillation Mode* (Academic, New York, 1997).
- Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D* **55**, 1690–1695 (1999).
- Collaborative Computational Project No. 4, The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Tervilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Tervilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Francis, N. R., Sosinsky, G. E., Thomas, D. & DeRosier, D. J. Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270 (1994).
- Morgan, D. G., Owen, C., Melanson, L. A. & DeRosier, D. J. Structure of bacterial flagellar filaments at 11 Å resolution: packing of the alpha-helices. *J. Mol. Biol.* **249**, 88–110 (1995).
- Case, D. A., *et al.* AMBER7. (Univ. California, San Francisco, California, 2002).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. M. Macnab, who passed away suddenly in September 2003, for his invaluable discussion on the structure and function of the flagellar hook; the staff members of beamline ID29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble and beamline BL41XU at the 8 GeV Super Photon ring (SPring-8) in Harima for their help for the data collection; and F. Oosawa and S. Asakura for continuous support and encouragement.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.N. (keiichi@fbs.osaka-u.ac.jp). Atomic coordinates have been deposited in the Protein Data Bank under accession code 1WLG.

The binary progenitor of Tycho Brahe's 1572 supernova

Pilar Ruiz-Lapuente^{1,2}, Fernando Comeron³, Javier Méndez^{1,4}, Ramon Canal¹, Stephen J. Smartt⁵, Alexei V. Filippenko⁶, Robert L. Kurucz⁷, Ryan Chornock⁶, Ryan J. Foley⁶, Vallery Stanishev⁸ & Rodrigo Ibata⁹

¹Department of Astronomy, University of Barcelona, Martí i Franques 1, E-08028 Barcelona, Spain

²Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany

³European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany

⁴Isaac Newton Group, PO Box 321, Santa Cruz de La Palma, Tenerife, Canary Islands, E-38780 Spain

⁵Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

⁶Department of Astronomy, 601 Campbell Hall, University of California, Berkeley, California 94720-3411, USA

⁷Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

⁸Department of Physics, Stockholm University, AlbaNova University Center, SE-108 91 Stockholm, Sweden

⁹Observatoire de Strasbourg, 11, rue de l'Université, F-67000 Strasbourg, France

The brightness of type Ia supernovae, and their homogeneity as a class, makes them powerful tools in cosmology, yet little is known about the progenitor systems of these explosions. They are thought to arise when a white dwarf accretes matter from a companion star, is compressed and undergoes a thermonuclear explosion^{1–3}. Unless the companion star is another white dwarf (in which case it should be destroyed by the mass-transfer process itself), it should survive and show distinguishing properties. Tycho's supernova^{4,5} is one of only two type Ia supernovae observed in our Galaxy, and so provides an opportunity to address observationally the identification of the surviving companion. Here we report a survey of the central region of its remnant, around the position of the explosion, which excludes red giants as the mass donor of the exploding white dwarf. We found a type G0–G2 star, similar to our Sun in surface temperature and luminosity (but lower surface gravity), moving at more than three times the mean velocity of the stars at that distance, which appears to be the surviving companion of the supernova.

Tycho Brahe's supernova (that is, SN 1572) is one of the only two supernovae observed in our Galaxy that are thought to have been of type Ia (the other having been SN 1006) as revealed by the light curve, radio emission and X-ray spectra^{4–7}.

The field that contained Tycho's supernova, relatively devoid of background stars, is favourable for searching for any surviving companion. With a Galactic latitude $b = +1.4^\circ$, Tycho's supernova lies 59–78 pc above the Galactic plane. The stars in that direction show a consistent pattern of radial velocities with a mean value of -30 km s^{-1} at 3 kpc. The predictions of how the companion star would look after the impact of the supernova ejecta, if there is any companion, depend on what the star actually is. The star could be in any evolutionary stage before the explosion: main sequence, sub-giant or red giant^{1–3}. The most salient feature of the surviving companion star should be peculiar velocities with respect to the average motion of the other stars at the same location in the Galaxy (mainly due to disruption of the binary)⁸, detectable through radial-velocity measurements, and perhaps also signs of the impact of the supernova ejecta. The latter can be twofold. First, mass should have been stripped from the companion and thermal energy injected into it, possibly leading to expansion of the stellar envelope that would make the star have a lower surface gravity. Second, depending on the

interaction with the ejected material, the surface of the star could be contaminated by the slowest-moving ejecta (made of Fe and Ni isotopes). If the companion's stellar envelope is radiative, such a contamination could be detectable through abundance measurements. Therefore, the observations have been designed along these lines. The star most likely to have been the mass donor of SN 1572 has to show a multiple coincidence: being at the distance of SN 1572, it has to show an unusual radial velocity in comparison to the stars at the same location (much above the velocity dispersion for its spectral type), and have stellar parameters consistent with being struck by the supernova explosion. It should also lie near the remnant centre (that is, within our search radius).

The distance to SN 1572 inferred from the expansion of the radio shell and by other methods lies around 3 kpc ($2.83 \pm 0.79 \text{ kpc}$)⁹. Such a distance, and the light-curve shape of SN 1572, are consistent with it being a normal type Ia supernova in luminosity, like those commonly found in cosmological searches⁹. Given the age of the supernova remnant (SNR; just 432 yr) and the lower limit to its distance, any possible companion, even if it moved at a speed of 300 km s^{-1} , could not be farther than 0.15 arcmin (9.1 arcsec) from its position at the time of the explosion^{8,10}. But the search radius

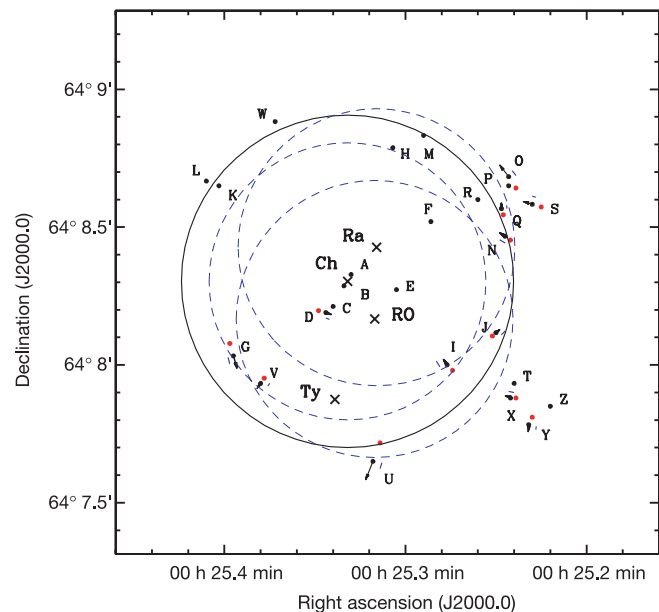


Figure 1 Positions and proper motions of stars. Positions are compared with three centres: the Chandra (Ch) and ROSAT (RO) geometrical centres of the X-ray emission, and that of the radio emission (Ra). Dashed lines indicate circles of 0.5 arcmin around those centres. The supernova position reconstructed from Tycho Brahe's measurements (Ty) is also shown, though merely for its historical interest²¹. The radius of the remnant is about 4 arcmin and the SNR is quite spherically symmetric, with a fairly good coincidence between radio and continuum X-ray emission^{8,22,23}. However, there is a 0.56 arcmin displacement along the east–west axis between the radio emission and the high-energy continuum in the 4.5–5.8 keV band observed by XMM–Newton in the position of the western rim²³ (Supplementary Note 1). Such asymmetry amounts to a 14% offset along the east–west axis. In SNRs from core-collapse supernovae (type II supernovae), up to a 15% discrepancy between the location of the compact object and the geometric centre is found in the most symmetric cases²⁴. On the basis of the above considerations, in our search we cover 15% of the innermost radius (0.65 arcmin) of the SNR around the Chandra centre of SN 1572. The companion star, if there is any, is unlikely to be outside this area (solid line). The proper motions of the stars measured from HST WFPC2 images are represented by arrows, their lengths indicating the total displacements between AD 1572 and present. Error bars are shown by parallel segments. Red circles are the extrapolated positions of the stars back to AD 1572. Star Tycho G displays a high proper motion, corresponding to the highest tangential velocity in the field, as both stars U and O are at much shorter distances (see Supplementary Methods).

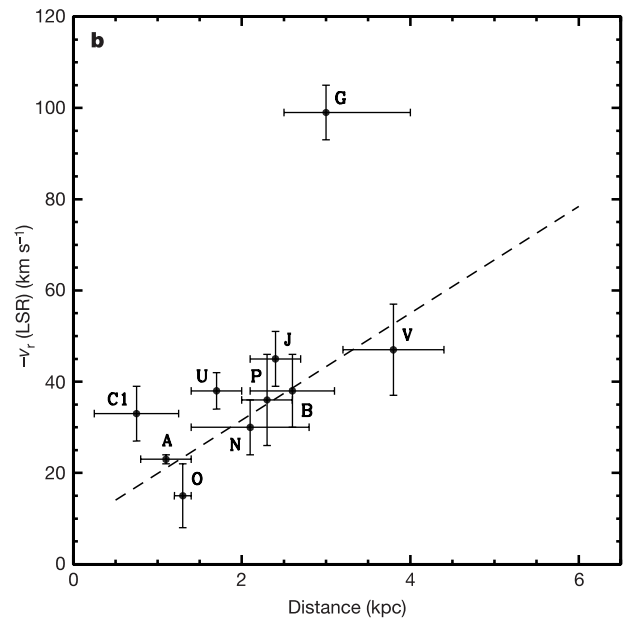
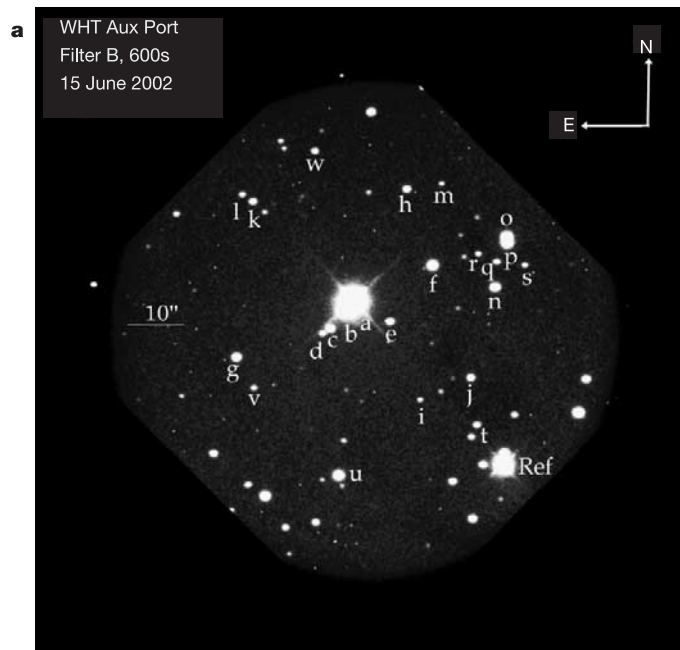


Figure 2 The SN 1572 field and radial velocity of the stars. **a**, Image from the Auxiliary Port at the William Herschel Telescope. It confirms the relative emptiness of the field. The search area (see also Fig. 1 bold circle) covers a radius of 0.65 arcmin around RA = 00 h 25 min 19.9 s, dec. = 64° 08' 18.2" (J2000) (the Chandra geometrical centre of X-ray emission) with repeated photometric and spectroscopic observations of the included stars at various epochs to check for variability and exclude binarity. Additional stars have been observed outside the 0.65 arcmin radius area and are visible in this field (whose diameter is 1.8 arcmin). For a remnant distance $d = 3.0$ kpc and a visual extinction $A_V = 1.7\text{--}2.0$ mag toward the candidate stars (see Supplementary Table 3), our search limit down to an apparent visual magnitude $V = 22$ implies that the survey must have detected all main-sequence stars of spectral types earlier than K6, plus all subgiant, giant

and supergiant stars within the corresponding cone. At that distance and with such extinction, the Sun would shine as a $V = 18.9$ mag star. **b**, Radial velocity (in the local standard of rest, LSR) versus distance for the subsample of stars closer than 6.5 kpc (the other stars are at a distance well beyond the SNR). We are looking outward along the Galactic plane, and the dashed line shows the approximate relationship for the stars in the direction of Tycho given by the expression $v_r = -v_\odot \cos(l - l_\odot) + A \sin(2l)$, where l and $l_\odot$ are the respective Galactic longitudes of Tycho and the solar apex, $v_\odot$ is the Sun's velocity in the LSR, and A is Oort's constant¹⁸. We include two field stars (stars O and U) that are slightly away from the search area (at $>15\%$ of the radius of the SNR) but at distances in the range 2–4 kpc as well. (Star names are labelled lower case in **a** for clarity.)

significantly expands owing to the uncertainty in the derived centre of the SNR (see Fig. 1).

We have analysed the stars within a circle of 0.65 arcmin radius, centred on the Chandra X-ray Observatory coordinates for the centre of the SNR, up to an apparent visual magnitude $V = 22$ (Figs 1 and 2, Table 1, and Supplementary Tables 1–3).

All but one of the stars found are either main-sequence stars (luminosity class V) with spectral types A4–K3 or giant stars (luminosity class III) with spectral types G0–K3.

Red-giant stars are possible companions of type Ia supernovae. Masses in the range 0.9–1.5 solar masses ($0.9\text{--}1.5 M_\odot$) would be the most favourable cases¹¹. Red giants are well represented in the sample, but none of them passes the tests for being a viable candidate. They are at distances incompatible with that of the supernova. The only giant relatively close to the distance of SN 1572 is Tycho A (Fig. 3), but it is closer than SN 1572 and shows no peculiarities in velocity, spectral type, or metallicity. Main-sequence stars are also viable companions of type Ia supernovae. Close binaries with 2 to $3.5 M_\odot$ main-sequence or subgiant companions have indeed been suggested as one class of systems able to produce type Ia supernovae¹². Among systems containing a main-sequence star, recurrent novae have been pointed out as possible progenitors¹³. Stripping of mass from the impact of the ejecta on this type of companion is also expected^{8,14}. Another consequence of the impact should be to puff up the star and dramatically increase its luminosity. The size and luminosity would later return to their equilibrium values for a star with the new decreased mass^{8,14,15}. Peculiar velocities should be highest ($200\text{--}300 \text{ km s}^{-1}$) in the case of main-sequence companions (orbital separations at the time of

Table 1 Characteristics of the supernova companion candidates

Star	θ (arcsec)	Spec. type and lum. class	T_{eff} (K)	$\log g$ (c.g.s.)	$E(B-V)$ (mag)	d (kpc)
Tycho A	1.6	K0–K1 III	4750	$2.5^{+0.5}_{-0.5}$	$0.55^{+0.05}_{-0.05}$	$1.1^{+0.3}_{-0.3}$
Tycho B	1.5	A8–A9 V	7500	$4.5^{+0.5}_{-0.5}$	$0.60^{+0.05}_{-0.05}$	$2.6^{+0.5}_{-0.5}$
Tycho C1	6.5	K7 V	4000	$4.5^{+0.5}_{-0.5}$	$0.5^{+0.1}_{-0.1}$	$0.75^{+0.5}_{-0.5}$
Tycho C2	6.5	F9 III	6000	$2.0^{+0.5}_{-0.5}$	$0.6^{+0.1}_{-0.1}$	>20
Tycho D	8.4	M1 V	3750	$4.5^{+0.5}_{-0.5}$	$0.6^{+0.3}_{-0.3}$	$0.8^{+0.2}_{-0.2}$
Tycho E	10.6	K2–K3 III	4250	$2.0^{+0.5}_{-0.5}$	$0.60^{+0.10}_{-0.10}$	>20
Tycho F	22.2	F9 III	6000	$2.0^{+0.5}_{-0.5}$	$0.54^{+0.22}_{-0.22}$	>10
Tycho G	29.7	G2 IV	5750	$3.5^{+0.5}_{-0.5}$	$0.60^{+0.05}_{-0.05}$	$3.0^{+1.0}_{-0.5}$
Tycho H	30.0	G7 III	5000	$3.0^{+0.05}_{-0.05}$	$0.60^{+0.09}_{-0.09}$	>13
Tycho J	33.9	K1 V	5000	$4.5^{+0.5}_{-0.5}$	$0.58^{+0.12}_{-0.11}$	$2.4^{+0.3}_{-0.2}$
Tycho K	35.0	F9 III	6000	$2.0^{+0.5}_{-0.5}$	$0.60^{+0.10}_{-0.10}$	>10
Tycho N	35.4	G0 V	6000	$4.5^{+0.5}_{-0.5}$	$0.62^{+0.08}_{-0.07}$	$2.1^{+0.7}_{-0.7}$
Tycho V	29.2	K3 V	4750	$4.5^{+0.5}_{-0.5}$	$0.60^{+0.10}_{-0.10}$	$3.8^{+0.6}_{-0.6}$

Supernova companion candidates within the search radius and limiting magnitude. Angular distances θ are from the Chandra X-ray geometrical centre, located at RA = 00 h 25 min 19.9 s, dec. = 64° 08' 18.2" (J2000). Synthetic spectra, under the assumption of local thermodynamic equilibrium (LTE), are fitted to the observed ones using the grids of model atmospheres and the atomic data of Kurucz²⁶, with the Uppsala Synthetic Spectrum Package²⁷. This determines the atmospheric parameters effective temperature T_{eff} and surface gravity g . Intrinsic colours and absolute visual magnitudes are deduced from the relationships between spectral type and colour and between spectral type and absolute magnitude for the different luminosity classes²⁸. Comparison with our photometric $B-V$ measurements (see Supplementary Table 3) yields the reddening $E(B-V)$, from which the visual extinction A_V and the corrected apparent visual magnitude V_0 are calculated. Comparison with the absolute visual magnitude then gives the distance d . Uncertainties in T_{eff} are 250 K. Tycho J is a binary of main-sequence stars with masses in the range $0.80\text{--}0.85 M_\odot$ and quite similar atmospheric parameters. Tycho C is found in HST images as being two stars (C1 and C2) 0.25 arcsec apart. Modelling of the composite spectrum and the HST magnitudes of the stars show that they do not constitute a physical binary; the hot fainter component C2 is at larger distance than C1. Within this list of stars, D, G, N and V have proper motions along Galactic longitude and latitude of $\mu_l = -3.23$, $\mu_b = -0.58 \pm 0.66$ (same error in both coordinates, units in mas yr^{-1}) for star D, $\mu_l = -2.60$, $\mu_b = -6.11 \pm 1.34$ for star G, $\mu_l = 3.23$, $\mu_b = 1.45 \pm 1.15$ for star N, and $\mu_l = 1.61$, $\mu_b = -2.85 \pm 0.78$ for star V.

explosion are shortest), but the measured values of radial velocity (v_r) for the main-sequence stars observed are not particularly high. The surface abundances are compatible with solar values. Other main-sequence stars (see Fig. 1, Table 1 and Supplementary Table 3) are found at wider separations from the geometrical centre, but they have v_r values within the range corresponding to their respective distances (see Supplementary Discussion).

We have found a subgiant star ('Tycho G') with lower surface gravity than that of main-sequence stars but higher surface gravity than that of red giants, which moves fast in comparison to the mean radial velocities of stars around it, and fits well the expectations for distance, reddening and velocity. Comparison of the Tycho G spectrum covering a wide wavelength range (3,180–9,400 Å) with templates¹⁶, after dereddening by $E(B - V) \approx 0.6$ mag, gives a best fit for an effective temperature $T_{\text{eff}} = 5,750$ K, a surface gravity $\log g$ between 4.0 and 3.0, and solar metallicity, which is confirmed by model fitting to high-resolution spectra in selected wavelength ranges (see Fig. 3, and Supplementary Fig. 1). For the spectral type found (G0–G2) and being a slightly evolved star (surface gravity not much below the main-sequence value), the mass should be about solar ($M \approx 1M_{\odot}$) and thus the radius, for the range of

surface gravities above, should be $R \approx 1\text{--}3R_{\odot}$, which translates (via our photometric data) into a distance $d \approx 2.5\text{--}4.0$ kpc. This companion could have been a main-sequence star or a subgiant before the explosion. While main-sequence companions might no longer look like ordinary main-sequence stars after the explosion of the type Ia supernova (and they might resemble subgiants, their envelopes having expanded after the supernova impact), subgiants would remain subgiants of lower surface gravity^{9,10,14,15}.

Stars at distances $d \approx 2\text{--}4$ kpc, in that direction, are moving at average radial velocity¹⁷ $v_r \approx -20$ to -40 km s⁻¹ (in the Local Standard of Rest), with a ~ 20 km s⁻¹ velocity dispersion^{18,19}. Tycho G moves at -108 ± 6 km s⁻¹ (heliocentric) in the radial direction. The deviation of Tycho G from the average thus exceeds by a factor of 3 the velocity dispersion of its stellar type. It has a 0.3% probability of having that characteristic and being unrelated to the explosion (that is, it is a 3σ outlier). In contrast, all other stars with distances compatible with that of SN 1572 have radial velocities within the velocity dispersion as compared with the average of all stars at the same location in the Galaxy. We studied through detailed proper motion measurements on the Hubble Space Telescope WFPC2 images²⁰ whether Tycho G has a high tangential velocity

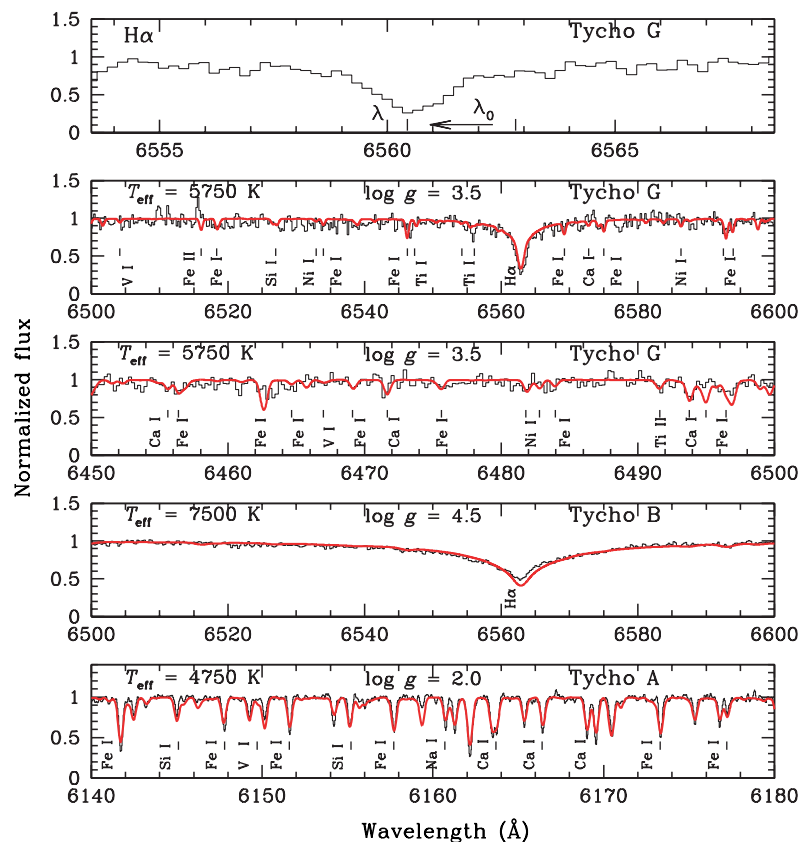


Figure 3 Model fits to observed spectra. Model atmosphere parameters are those listed in Table 1, and chemical abundances are solar. They are shown here for our candidate star for the companion of SN 1572 (Tycho G) and the red giant (Tycho A) and main-sequence star (Tycho B) nearest to the distance of SN 1572 and to the SNR X-ray centre. Identifications of the most significant metal lines are given. We have not detected significant spectroscopic anomalies, either here or in the whole sample, and most spectra are well reproduced assuming solar abundances²⁵. Thin lines correspond to the observations and thicker lines to the synthetic spectra. Spectra were obtained at the William Herschel Telescope (WHT) with UES and ISIS. Tycho A (bottom panel) is the closest red giant in the sample. It is a K0 III star, and its mass should be typically $M \approx 3M_{\odot}$ ($M_{\odot}$ stands for the mass of the Sun). Here, Tycho A is ruled out simply on the basis of having too short a distance. All the other red giants are located well beyond Tycho's remnant, and

therefore also ruled out (see Supplementary Discussion). The A8/A9 star Tycho B (second panel from bottom) has $M \approx 1.5M_{\odot}$, which would fall within the appropriate range for main-sequence type Ia supernova companions, as it would have been massive enough to transfer the required amount of mass to the white dwarf. The entirely normal atmospheric parameters, however, strongly argue against any such event in the star's recent past. The low radial velocity reinforces this conclusion. Tycho G (three upper panels). The second and third spectra from the top show computed spectra compared with observed spectra obtained at the WHT with ISIS. The upper panel shows the observed spectrum near H α . This line is blueshifted, implying a peculiar radial velocity exceeding about 3 times the velocity dispersion for its stellar type. This star does not belong to the halo population (Supplementary Fig. 1).

as well (see Supplementary Table 2 and Supplementary Methods). Tycho G has significant proper motion toward lower Galactic latitude: $\mu_b = -6.11 \pm 1.34 \text{ mas yr}^{-1}$ (the proper motion along longitude is small, $\mu_l = -2.6 \pm 1.34 \text{ mas yr}^{-1}$). The proper motion in Galactic latitude implies that this star is an outlier in proper motion as well, with a derived tangential velocity of $94 \pm 27 \text{ km s}^{-1}$ (a 24 km s^{-1} systematic error was added, resulting from a 1.7 mas yr^{-1} uncertainty in the reference frame solution of the images). The other stars do not show such coincidence in distance and high tangential velocity. The modulus of the velocity vector has a value of 136 km s^{-1} , which is a factor of over 3 larger than the mean velocity value at 3 kpc.

If Tycho G is the companion star as suggested by its kinematics, the explosion centre should have been 2.6 arcsec north of the current location of this star on the basis of its velocity. The peculiar velocity would correspond to the peculiar velocities expected from the disruption of a white dwarf plus subgiant/main-sequence system^{9,10} of roughly a solar mass. The system would have resembled the recurrent nova U Scorpii (see Supplementary Note 2). The excess velocity corresponds to a period of about 2–7 days, for a system made of a white dwarf close to the Chandrasekhar mass plus a companion of roughly a solar mass at the moment of the explosion.

Several paths lead to this star as the likely donor star of SN 1572: its high peculiar velocity (both radial and tangential velocities), the distance in the range of SN 1572, and its type, which fits the post-explosion profile of a type Ia supernova companion, as the position of this star in the Hertzsprung–Russell diagram is also untypical for a standard subgiant. The lower limit to the metallicity obtained from the spectral fits is $[M/H] > -0.5$ (see Fig 3 and Supplementary Fig. 1), which excludes its belonging to the Galactic halo population as an alternative explanation of its high velocity. Spectra taken at five different epochs also exclude its being a single-lined spectroscopic binary. If our candidate is the companion star, its overall characteristics imply that the supernova explosion affected the companion mainly through the kinematics. Our search for the binary companion of Tycho's supernova has excluded giant stars. It has also shown the absence of blue or highly luminous objects as post-explosion companion stars. A star very similar to the Sun but of a slightly more evolved type is here suggested as the likely mass donor that triggered the explosion of SN 1572. That would connect the explosion to the family of cataclysmic variables. □

Received 14 June; accepted 8 September 2004; doi:10.1038/nature03006.

- Branch, D., Livio, M., Yungelson, L. R., Boffi, F. & Baron, E. In search of the progenitors of Type Ia Supernovae. *Publ. Astron. Soc. Pacif.* **107**, 1019–1029 (1995).
- Ruiz-Lapuente, P., Canal, R. & Burkert, A. in *Thermonuclear Supernovae* (eds Ruiz-Lapuente, P., Canal, R. & Isern, J.) 205–230 (Kluwer Academic, Dordrecht, 1997).
- Livio, M. in *Type Ia Supernovae: Theory and Cosmology* (eds Niemeyer, J. C. & Truran, J. W.) 33–48 (Cambridge Univ. Press, Cambridge, 2000).
- Brahe, T. *Astronomiae Instauratae Progymnasmata*. In *Opera Omnia* Vol. 2 (ed. Dreyer, I. L. E.) 307 (Swets & Zeitlinger, Amsterdam, 1972) (1603).
- Baade, W. B. Cassiopeiae as a supernova of Type I. *Astrophys. J.* **102**, 309–317 (1945).
- Baldwin, J. E. & Edge, D. O. Radio emission from the remnants of the supernovae of 1572 and 1604. *Observatory* **77**, 139–143 (1957).
- Hughes, J. P. et al. ASCA observations of the Large Magellanic Cloud supernova remnant sample: Typing supernovae from their remnants. *Astrophys. J.* **444**, L81–L84 (1995).
- Canal, R., Méndez, J. & Ruiz-Lapuente, P. Identification of the companion stars of Type Ia supernovae. *Astrophys. J.* **550**, L53–L56 (2001).
- Ruiz-Lapuente, P. Tycho Brahe's supernova: light from centuries past. *Astrophys. J.* **612**, 357–363 (2004).
- Ruiz-Lapuente, P. et al. in *From Twilight to Highlight: the Physics of Supernovae* (eds Hillebrandt, W. & Leibundgut, B.) 140–147 (Springer, Berlin, 2003).
- Hachisu, I., Kato, M. & Nomoto, K. A new model for progenitor systems of Type Ia supernovae. *Astrophys. J.* **470**, L97–L100 (1996).
- Li, X.-D. & van den Heuvel, E. P. J. Evolution of white dwarf binaries: supersoft X-ray sources and progenitors of type Ia supernovae. *Astron. Astrophys.* **322**, L9–L12 (1997).
- Livio, M. & Truran, J. W. Type I supernovae and accretion-induced collapses from cataclysmic variables? *Astrophys. J.* **389**, 695–703 (1992).
- Marietta, E., Burrows, A. & Fryxell, B. Type Ia supernova explosions in binary systems: the impact on the secondary star and its consequences. *Astrophys. J. Suppl.* **128**, 615–650 (2000).
- Podsiadlowski, P. On the evolution and appearance of a surviving companion after a Type Ia supernova explosion. Preprint at (<http://arxiv.org/astro-ph/0303660>) (2003).

- Lejeune, Th., Cuisinier, F. & Buser, R. A standard stellar library for evolutionary synthesis. *Astron. Astrophys. Suppl. Ser.* **125**, 229–246 (1997).
- Brand, J. & Blitz, L. The velocity field of the outer Galaxy. *Astron. Astrophys.* **275**, 67–90 (1993).
- Binney, J. & Merrifield, M. *Galactic Astronomy* (Princeton Univ. Press, Princeton, 1998).
- Dehnen, W. & Binney, J. Local stellar kinematics from *Hipparcos* data. *Mon. Not. R. Astron. Soc.* **298**, 387–394 (1998).
- Ruiz-Lapuente, P. Probing the nature of Type Ia SNa through HST astrometry. (http://archive.stsci.edu/cgi-bin/proposal_search?mission=hst&id=9729) (2003).
- Ruiz-Lapuente, P. in *Proc. Conf. Astronomy as a Model for the Sciences in Early Modern Times, Munich, 21–23 March 2003* (eds Fritscher, B. & Kühne, A.) (Erwin Rauner, Augsburg, in the press).
- Hughes, J. P. The expansion of the X-ray remnant of Tycho's supernova (SN 1572). *Astrophys. J.* **545**, L53–L56 (2000).
- Decourchelle, A. et al. XMM-Newton observation of the Tycho supernova remnant. *Astron. Astrophys.* **365**, L218–L224 (2001).
- (<http://ccx.harvard.edu/cda/>)
- Anders, E. & Grevesse, N. Abundances of the elements—meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Kurucz, R. L. *ATLAS9 Stellar Atmospheres Programs, Grids of Model Atmospheres and Line Data [CD-ROM]*. (Smithsonian Astrophysical Observatory, Cambridge, MA, 1993).
- Gustafsson, B., Bell, R. A., Eriksson, K. & Nordlund, Å. A grid of model atmospheres for metal-deficient giant stars. *Astron. Astrophys.* **42**, 407–432 (1975).
- Schmidt-Kaler, Th. in *Landolt-Börnstein, New Ser. VI*, Vol. 2b (ed. Hellwege, K.-H.) 1–34 (Springer, Berlin, 1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements P.R.L. thanks C. Ruiz Ogara for giving her the spirit to complete this survey. We thank the support staff at the European Northern Observatory at La Palma for their assistance throughout this project, as well as the support staff at the W. M. Keck Observatory and NASA/ESA Hubble Space Telescope. We express our special gratitude to C. Abia, F. Figueras, C. Guirao, R. Mignani and J. Torra for diverse consultations. This work has been supported by DURSI, DGCYT (to P.R.L., J.M. and R.C.), PPARC (to J.M. and S.J.S.) and by NSF (to A.V.E., R.C., R.J.F. and R.L.K.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests should be addressed to P.R.L. (pilar@am.ub.es).

Recent ice-rich deposits formed at high latitudes on Mars by sublimation of unstable equatorial ice during low obliquity

Benjamin Levrard^{1*}, François Forget², Franck Montmessin³ & Jacques Laskar¹

¹Astronomie et Systèmes Dynamiques, IMC-CNRS UMR8028, 77 Avenue Denfert-Rochereau, 75014 Paris, France

²Laboratoire de Météorologie Dynamique, Université Paris VI, 4 Place Jussieu, 75005 Paris, France

³Space Science Division MS 245-3, NASA/Ames Research Center, Moffett Field, California 94035, USA

* Present address: Laboratoire de Planétologie et téledétection, UMR 5570, UCBL-ENS Lyon, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne, France

Observations from the gamma-ray spectrometer instrument suite on the Mars Odyssey spacecraft have been interpreted as indicating the presence of vast reservoirs of near-surface ice in high latitudes of both martian hemispheres^{1–5}. Ice concentrations are estimated to range from 70 per cent at 60° latitude to 100 per cent near the poles, possibly overlain by a few centimetres of ice-free material in most places⁴. This result is supported by morphological evidence of metres-thick layered deposits that are rich in water-ice^{6–9} and periglacial-like features^{10,11} found only at high latitudes. Diffusive exchange of water between the pore space of the regolith and the atmosphere has been proposed to explain this distribution¹², but such a degree of concentration is

difficult to accommodate with such processes^{9,13,14}. Alternatively, there are suggestions that ice-rich deposits form by transport of ice from polar reservoirs and direct redeposition in high latitudes during periods of higher obliquity^{9,13}, but these results have been difficult to reproduce with other models. Here we propose instead that, during periods of low obliquity (less than 25°), high-latitude ice deposits form in both hemispheres by direct deposition of ice, as a result of sublimation from an equatorial ice reservoir that formed earlier, during a prolonged high-obliquity excursion. Using the ice accumulation rates estimated from global climate model simulations we show that, over the past ten million years, large variations of Mars' obliquity have allowed the formation of such metres-thick, sedimentary layered deposits in high latitude and polar regions.

On Earth, quasi-periodic variations in orbital (eccentricity, longitude of perihelion from the moving equinox L_p) and spin axis parameters are implicated in significant climatic changes in the past million years (Myr)¹⁵. Oscillations between glacial and interglacial periods are characterized by a transfer of large amounts of water between polar ice sheets and oceans. Similarly, astronomical forcing and consequent changes in seasonal and global insolation are presumed to cause large-scale redistribution and cycling of volatiles at Mars' surface over timescales of $\sim 10^4$ – 10^6 years. In particular, previous simplified and full three-dimensional (3D) climate models have predicted that at much higher obliquities ($>40^\circ$), polar ice may sublime rapidly and be redeposited in the tropics^{13,16–19}.

The variations of obliquity and orbital parameters of Mars are strongly chaotic^{20,21}. Nevertheless, our knowledge of the present rotational state of Mars is sufficient to give a reliable solution of their evolution for the past 10 Myr (refs 22, 23). Over this interval, Mars' obliquity is characterized by a marked transition around 4 Myr ago between a high-mean-obliquity regime of $\sim 35 \pm 10^\circ$ and a more recent low-obliquity regime of $\sim 25 \pm 10^\circ$ (Fig. 1), while its eccentricity has varied between 0 and ~ 0.12 with a dominant ~ 2.4 -Myr modulating period. Here, we use the martian global climate model (GCM)^{24,25} of the Laboratoire de Météorologie Dynamique (LMD) to investigate the evolution of surface ice deposits across the large obliquity changes of this transition. We used a horizontal resolution of 7.5° in longitude and 5.625° in latitude and 25 vertical levels. The model includes a full description of exchange between surface ice and atmospheric water, transport and turbulent mixing of water in the atmosphere and cloud formation^{25,26}. The radiative effects of water vapour and clouds as well as the exchanges of water vapour with the subsurface are not included. The surface albedo is set to 0.4 when an ice layer thicker than $5 \mu\text{m}$ is present, enabling an ice-albedo feedback process. The surface thermal inertia is not modified, however. Control simulations of seasonal water cycle for the present-day orbital parameters provide latitudinal distributions of atmospheric vapour and clouds in very good agreement with TES spectrometer observations^{26,27}.

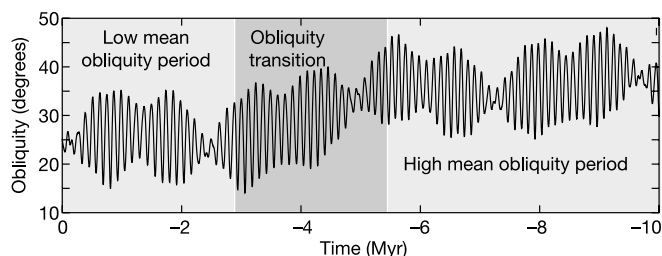


Figure 1 Evolution of the martian obliquity in the past 10 Myr. Characteristic periods are indicated: high- and low-mean obliquity periods in pale grey, obliquity transition in dark grey; after Laskar *et al.*²³. The main obliquity periodicity is about 120,000 earth years.

In a first set of simulations, we investigated the global stability of the northern ice cap for obliquity values ranging from the present value (25.19°) to 45° in 5° steps. In each case, the model is spun up from dry initial conditions with a northern residual ice cap as the only initial water source and then run until the atmosphere comes to an interannually repeatable state. The cap is 'unstable' (it undergoes a net loss) when the water lost in summer is not transported back during the rest of the year. To first order, we found that the net amount of water sublimated from the cap (per unit surface area) and transported to nonpolar regions depends mainly on the summer polar insolation, which is a sensitive function of both obliquity and orbital parameters.

For circular orbits, the cap is unstable for obliquities higher than 35° and the corresponding annual loss rates are 6.5, 27.8 and 65.0 mm per martian year for obliquities ranging from 35 to 45° . However, for an orbit with the current Mars eccentricity (0.0934), and with a perihelion coincident with the northern summer solstice ($L_p = 90^\circ$), thus maximizing the summer polar insolation, the northern cap already becomes unstable at 30° obliquity. Annual water loss rates then reach 5.0, 26.5, 80.2 and 218 mm per martian year for obliquities ranging from 30 to 45° , respectively. Conversely, when the northern summer solstice corresponds to the aphelion ($L_p = 270^\circ$), obliquities of 40° and higher are required to obtain an unstable cap and the annual loss rates are 7.2 and 21.6 mm yr^{-1} for 40° and 45° values, respectively. Thus, in our model, there is a critical obliquity between 35 and 40° , above which the stability of the north polar cap is lost whatever the values of the orbital parameters, in agreement with ref. 13.

In all these cases, the LMD model predicts substantial ice deposition in the tropics, in agreement with previous GCM studies^{13,18}. However, unlike in ref. 13 no accumulation of stable ice was ever found in the mid- and high latitudes for any obliquities when assuming a polar source. The NASA Ames GCM also does not reproduce this behaviour (R. M. Haberle, personal communication).

In our simulations, when the polar cap is unstable, ice accumulation occurs in the high topography areas of Tharsis Montes and Olympus Mons where the saturation state of the atmosphere becomes the highest (Fig. 2). Ice precipitation is favoured by adiabatic cooling and upwelling of atmospheric water on the windward slopes. For a typical circular orbit, net ice accumulation rates may reach $\sim 15 \text{ mm}$ per martian year at 35° , 30 mm yr^{-1} at 40°

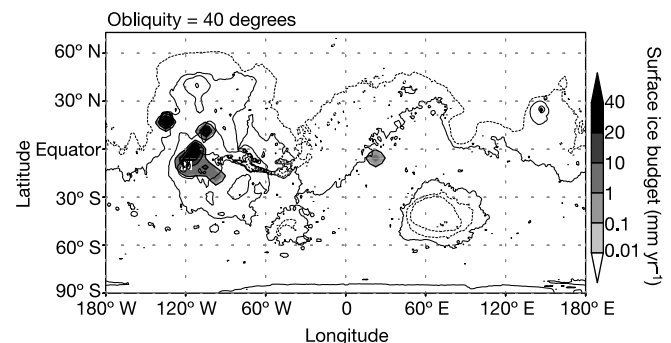


Figure 2 Surface water ice budget (in millimetres per martian year) for a 40° obliquity during simulation year 11, with superimposed MOLA topography. The eccentricity is set to zero. Main accumulation regions are located on Tharsis Rise (around Arsia, Pavonis and Ascraeus Montes) and Olympus Mons. Minor accumulation ($\sim 1 \text{ mm}$ per martian year) occurs near the equatorial Schiaparelli basin impact. For other obliquities and/or orbital parameters leading to a vanishing northern water ice cap, the locations of stable surface ice are similar and also concentrated on the Tharsis area and Olympus Mons, but with local variations of ice accumulation rates. Climate simulations performed with higher resolution (3.75° in longitude and 5.625° in latitude) result in more localized ice sheets on Tharsis Montes and Olympus Mons.

and 60 mm yr^{-1} at 45° on some of Tharsis Montes. We thus believe that, during the mean high-obliquity regime ($\sim 5\text{--}10 \text{ Myr}$ ago), and unless a dust lag inhibited polar sublimation, a few obliquity cycles may have been sufficient to empty the polar reservoirs and form massive glaciers $\sim 3 \text{ km}$ thick or more on the Tharsis Montes and Olympus Mons, depending on the amount of polar water available. This could explain the presence of fan-shaped deposits interpreted as remnant traces of 'cold-based' glaciers on their western flanks^{28,29}.

When the obliquity dropped below the critical value, the equatorial ice reservoirs began to sublime and lose water, leading to the redistribution of ice to other areas on the surface of Mars. To investigate where the ice may have gone under these circumstances, we performed climate simulations for obliquities 30 , 25.19 , 20 and 15° , all starting with an equatorial ice reservoir as the only source of water on the surface. We tested various sizes and locations of these reservoirs and various initial atmospheric water contents. These simulations yielded similar results. Below, we discuss results from simulations performed assuming an initial dry atmosphere and an ice reservoir on Tharsis Montes (Fig. 3), for a circular orbit.

The annual surface ice budgets obtained from these simulations (Fig. 3) show that the ice returns to the high-latitude and polar areas of both hemispheres, but that the extent of the ice distribution is sensitive to obliquity. At 30° obliquity, the ice accumulates only in the northern polar areas. Because the eccentricity here is set to zero, we attributed this asymmetry to biasing of the general circulation by the topography differences between the two hemispheres³⁰. At the current obliquity ($\sim 25.2^\circ$), ice covers most areas north of 65° N and begins to accumulate near the south pole. Finally, at 20° and 15° obliquities, large-scale ice accumulation is observed poleward of $\sim 60^\circ$ in both hemispheres with some excursions into the mid-latitude regions ($\sim 30\text{--}60^\circ \text{ S}$) of the Southern hemisphere. Despite

its zonal dependence, the boundaries of the global ice distribution exhibit significant coincidence with the near-surface ice distribution inferred from GRS data¹⁻⁵ and latitude-dependent ice-rich deposits from MOLA and MOC data²⁹, especially in the northern hemisphere.

The fact that the surface ice becomes stable and accumulates at subpolar latitudes for obliquity equal to or lower than today's when an equatorial source is present is an unavoidable consequence of the high water-vapour content of the atmosphere under such conditions. Currently on Mars at 25.19° obliquity, surface ice also accumulates in autumn, winter and early spring at $60\text{--}75^\circ$ latitude. However, this layer of ice always remains thinner than 0.5 mm on average (in our current Mars GCM simulation) and quickly sublimates after mid-spring. When an equatorial source is present, much more water vapour is transported to the high latitudes and the amount of ice that accumulates during autumn, winter and early spring can reach several millimetres above 65° N . This is more than the amount that can sublime in late spring ($\sim 1\text{--}2 \text{ mm}$). At obliquity lower than today's, the water content of the global atmosphere is about the same as at the current obliquity, but the late spring and summer ice sublimation is lower because of the reduced insolation.

Ice deposition rates in high and polar latitudes appear to be weakly sensitive to the obliquity value (Fig. 3). Above 65° N , these rates remain close to $\sim 2 \text{ mm}$ per martian year between 25 to 15° obliquity, and are slightly lower in the southern hemisphere. With such low rates, it is likely that a significant amount of dust was incorporated in the icy deposits if the atmosphere was as dusty as today's. Nevertheless, assuming a steady accumulation of 2 mm per martian year, we can estimate that much more than 10 m of ice-rich material should have accumulated during each obliquity cycle when

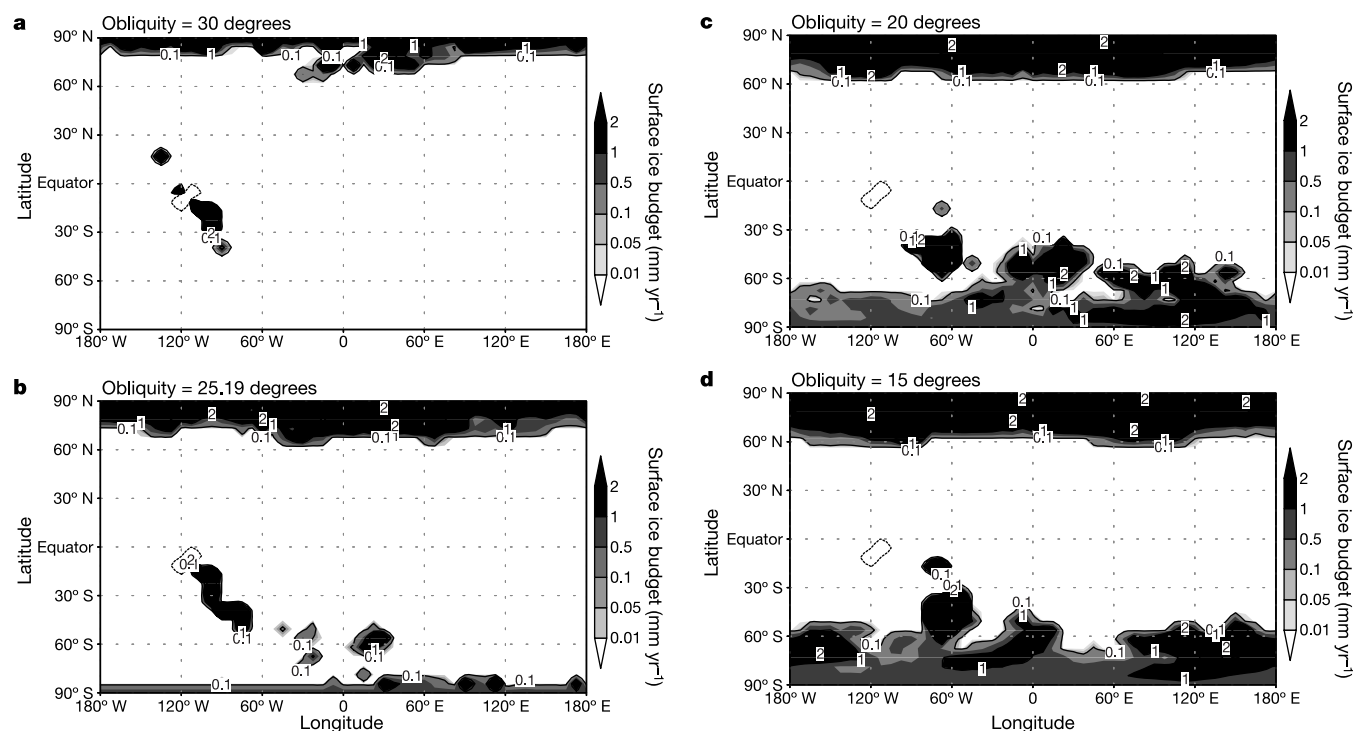


Figure 3 Surface water ice budget in mm per martian year after ten years of simulation for various obliquities, and with an example of equatorial ice reservoir situated around Arsia and Pavonis Montes. **a**, 30° ; **b**, present obliquity 25.19° ; **c**, 20° ; **d**, 15° . Its boundaries are indicated by a thick solid line. We assume dry initial conditions and the equatorial reservoir as the only water source. The eccentricity is set to zero. In that case, the critical obliquity for the northern polar cap stability is between 30 and 35° . At 25.19° and 30° obliquities,

local accumulation occurs in tropical zones around the equatorial source, which depends on the extent and the location of the source. At high latitudes, ice accumulation is more uniformly distributed in longitude and latitude in the northern than in the southern hemisphere. Typical accumulation rates ($\sim 2 \text{ mm}$ per martian year) are an order of magnitude lower than the ice loss rates at high obliquity, suggesting that pure polar ice deposits may not survive during the high mean obliquity regime ($\sim 5\text{--}10 \text{ Myr}$ ago).

the obliquity was below 25°, because each of these episodes lasted about 60,000 years. A significant part of this ice probably sublimed during the other half of the obliquity cycle with obliquity above 30°. However, as suggested by refs 13 and 14, the upper part of the ice layers sublimated, but probably left behind a dry slab layer that protected deeper ice deposits from further sublimation. If this is accurate, then each obliquity cycle created at least one distinct layer which may still be present below the surface of Mars.

However, this process supposes that an equatorial ice reservoir remained present during periods of low obliquity. To test this hypothesis, we analysed the ice loss rates of individual equatorial sources at low obliquity for initial ice distributions issued from high-obliquity simulation distributions (Fig. 2). We found minimal loss rates of ~1.0, 10, 15 and 20 mm per martian year at 30, 25.19, 20 and 15° obliquity, respectively. Integrating a simple deposition/sublimation history over the obliquity cycles of the obliquity transition shows that a 3-km-thick equatorial ice source may have survived throughout the whole transition until ~3 Myr ago.

Once the equatorial sources were exhausted, what finally happened to the high-latitude ice deposits? To address this issue, we performed simulations starting with the initial ice inventory of the previous simulations (as on Fig. 3) but without the equatorial source. We found that the edges of the mid- and high-latitude deposits become unstable and are redeposited poleward. Further simulations indicate that the surface ice slowly retreats to the pole. There again, it is likely that in reality only the upper ice sublimed and that a significant part of the layer remained under a protecting lag deposit, the one that we still see today at the surface of Mars.

Over longer timescales (~10⁹ years), the 'chaotic diffusion' of Mars' obliquity shows that periods of high obliquity (>40°) are statistically the most probable situation²³, suggesting that our mechanism could have acted throughout martian geological history. In this context, it is possible that the near-surface ice detected by Mars Odyssey is only the upper layer of a much deeper ice reservoir. Future MARSIS and SHARAD sounding investigations aboard Mars Express and the Mars Reconnaissance Orbiter may provide important constraints on these reservoirs. □

Received 8 April; accepted 17 September 2004; doi:10.1038/nature03055.

- Mitrofanov, I. *et al.* Maps of subsurface hydrogen from the high energy neutron detector, Mars Odyssey. *Science* **297**, 78–81 (2002).
- Boynton, W. V. *et al.* Distribution of hydrogen in the near surface of Mars: evidence for subsurface ice deposits. *Science* **297**, 81–85 (2002).
- Feldman, W. C. *et al.* Global distribution of neutrons from Mars. *Science* **297**, 75–78 (2002).
- Boynton, W. V. *et al.* Constraints on the distribution of hydrogen in the polar regions of Mars and implications for ice formation processes. *AGU Fall Meeting Abstr.* P32B-05 (2003).
- Feldman, W. C. *et al.* The global distribution of near-surface hydrogen on Mars. *J. Geophys. Res.* **109** (E6), doi:10.1029/2003JE002160 (2004).
- Mustard, J. F., Cooper, C. D. & Rifkin, M. K. Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* **412**, 411–414 (2001).
- Kreslavsky, M. A. & Head, J. W. Kilometre-scale roughness of Mars: Results from MOLA data analysis. *J. Geophys. Res.* **105**, 26695–26711 (2000).
- Kreslavsky, M. A. & Head, J. W. Nature and evolution of young latitude-dependent water-ice rich mantle. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL015392 (2002).
- Head, J. W., Mustard, J. F., Kreslavsky, M. A., Milliken, R. E. & Marchant, D. R. Recent ice ages on Mars. *Nature* **426**, 797–802 (2003).
- Mangold, N., Maurice, S., Feldman, W. C., Costard, F. & Forget, F. Geographical relationships between small scale polygons and ground ice distribution from neutron spectrometer on Mars. *Third Mars Polar Sci. Conf.* Abstr. 8043 (2003).
- Milliken, R. E. *et al.* Viscous flow features on the surface of Mars. *J. Geophys. Res.* **108** (E6), doi:10.1029/2002JE002005 (2003).
- Mellon, M. T., Feldman, W. C. & Prettyman, T. H. The presence and stability of ground ice in the Southern hemisphere of Mars. *Icarus* **169**, 324–340 (2004).
- Mischna, M. A., Richardson, M. I., Wilson, R. J. & McCleese, D. J. On the orbital forcing of Martian water and CO₂ cycles: A general circulation model study with simplified volatile schemes. *J. Geophys. Res.* **108** (E6), doi:10.1029/2003JE002051 (2003).
- Mischna, M. A., McCleese, D. J., Richardson, M. I., Vasavada, A. R. & Wilson, R. J. Volatile cycling and layering on Mars: Observations, theory and modeling. *6th Int. Mars Conf.* Abstr. 3145 (2003).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations of the Earth's Orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Jakosky, B. M. & Carr, M. H. Possible precipitation of ice at low latitudes of Mars during periods of high obliquity. *Nature* **315**, 559–561 (1985).
- Jakosky, B. M., Henderson, B. G. & Mellon, M. T. Chaotic obliquity and the nature of the Martian climate. *J. Geophys. Res.* **100**, 1579–1584 (1995).

- Haberle, R. M. *et al.* The Martian water cycle at high obliquity. *Lunar Planet. Sci. Conf.* XXXI Abstr. 1509 (2000).
- Richardson, M. I. & Wilson, R. J. Investigation of the nature and stability of the Martian seasonal water cycle with a general circulation model. *J. Geophys. Res.* **107** (E5), doi:10.1029/2001JE001536 (2002).
- Laskar, J. & Robutel, P. The chaotic obliquity of the planets. *Nature* **361**, 608–612 (1993).
- Touma, J. & Wisdom, J. The chaotic obliquity of Mars. *Science* **259**, 1294–1297 (1993).
- Laskar, J., Levrard, B. & Mustard, J. F. Orbital forcing of the martian polar layered deposits. *Nature* **419**, 375–377 (2002).
- Laskar, J. *et al.* Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* **170**, 343–364 (2004).
- Houde, F., Forget, F. & Talagrand, O. Meteorological variability and the annual surface pressure cycle on Mars. *J. Atmos. Sci.* **50**, 3625–3640 (1993).
- Forget, F. *et al.* Improved General Circulation Models of the Martian atmosphere from the surface to above 80 km. *J. Geophys. Res.* **104**, 24155–24176 (1999).
- Montmessin, F., Forget, F., Rannou, P., Cabane, M. & Haberle, R. M. Origin and role of water ice clouds in the Martian water cycle as inferred from a general circulation model. *J. Geophys. Res.* **109** (E10), doi:10.1029/2004JE002284 (2004).
- Smith, M. D. The annual cycle of water vapor on Mars as observed by the Thermal Emission Spectrometer. *J. Geophys. Res.* **107** (E6), doi:10.1029/2001JE001522 (2002).
- Head, J. W. & Marchant, D. R. Cold-based mountain glaciers on Mars: Western Arsia Mons. *Geology* **31**, 641–644 (2003).
- Shean, D. E., Head, J. W., Fastook, J. L. & Marchant, D. R. Tharsis Montes cold-based glaciers: Observations and constraints for modeling and preliminary results. *Lunar Planet. Sci. Conf.* XXXV Abstr. 1428 (2004).
- Richardson, M. I. & Wilson, R. J. A topographically forced asymmetry in the martian circulation and climate. *Nature* **416**, 298–301 (2002).

Acknowledgements We acknowledge R. M. Haberle and G. L. Tyler for useful discussions and comments. This work was supported by the CNRS-PNP programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.L. (blevraud@imcce.fr).

Continuous generation of single photons with controlled waveform in an ion-trap cavity system

Matthias Keller¹, Birgit Lange¹, Kazuhiro Hayasaka², Wolfgang Lange¹ & Herbert Walther^{1,3}

¹Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany

²National Institute of Information and Communications Technology, 588-2 Iwaoka, Nishi-ku, Kobe 651-2492, Japan

³Sektion Physik der Universität München, Am Coulombwall 1, 85748 Garching, Germany

The controlled production of single photons is of fundamental and practical interest; they represent the lowest excited quantum states of the radiation field, and have applications in quantum cryptography¹ and quantum information processing². Common approaches use the fluorescence of single ions³, single molecules^{4,5}, colour centres^{6,7} and semiconductor quantum dots^{8–12}. However, the lack of control over such irreversible emission processes precludes the use of these sources in applications (such as quantum networks¹³) that require coherent exchange of quantum states between atoms and photons. The necessary control may be achieved in principle in cavity quantum electrodynamics. Although this approach has been used for the production of single photons from atoms^{14–16}, such experiments are compromised by limited trapping times, fluctuating atom–field coupling and multi-atom effects. Here we demonstrate a single-photon source based on a strongly localized single ion in an optical cavity. The ion is optimally coupled to a well-defined field mode, resulting in the generation of single-photon pulses with precisely defined shape and timing. We have confirmed the

suppression of two-photon events up to the limit imposed by fluctuations in the rate of detector dark counts. The stream of emitted photons is uninterrupted over the storage time of the ion, as demonstrated by a measurement of photon correlations over 90 min.

Two fundamentally different approaches have been used to produce single-photon states. In the first case, the photons are generated spontaneously, in response to fast excitation, for example, in fluorescence of single emitters^{3–12}. Although remarkable improvements have been made in efficiency, two-photon suppression and spectral properties, these sources emit the photon in an irreversible process and thus cannot serve as a bidirectional interface between atomic and photonic quantum states, as required in a quantum network¹³. In contrast, in this Letter, we focus our attention on a system in which the photon is generated as the result of a coherent process. At present, the necessary conditions are provided exclusively in the framework of cavity quantum electrodynamics (cavity-QED), where a single atom is strongly coupled to the electromagnetic field mode into which the photon is emitted. Single-photon generation in this scheme has been proposed previously¹⁷ and demonstrated in recent experiments^{14,15}. The duration of photon emission was limited by the lifetime of the dipole trap used for storing the atom, the best values reaching a fraction of a second¹⁵. In addition, thermal motion of the atom produced variations of the atom-field coupling¹⁸, while the residual probability of two atoms being in the trap resulted in a considerable number of two-photon events¹⁵. We have overcome these limitations of single-photon generation with atoms by using a single tightly bound ion as the source of radiation, which is optimally coupled to the radiation field for many hours. In this way, we have realized a cavity-QED set-up, in which the coupling strength of a trapped ion and a quantized field-mode is deterministically controlled. This achievement is a key requirement for a range of applications in quantum information processing^{13,19–21}.

The experimental apparatus we use for trapping a single calcium

ion and localizing it in the field of an optical resonator has been described previously^{22,23} and is sketched in Fig. 1. Briefly, we use a linear radio-frequency ion trap to confine a single $^{40}\text{Ca}^+$ ion at the centre of an optical resonator, tuned to the $D_{3/2} \rightarrow P_{1/2}$ transition in calcium at a wavelength of 866 nm (see inset of Fig. 1). In a radial potential well of frequency 1.3 MHz, we have reached a root-mean-square (r.m.s.) spread of the ion's wavepacket of the order of 40 nm^{22,23}. This corresponds to a relative uncertainty of the coupling g of 2% at the maximum of the standing wave. The permanent and fully controllable coupling of a particle to the field with a well-defined strength is a unique property of ion-trap cavity-QED and provides the basis for the experiments reported here. The values of the coupling g and the cavity damping κ for a cavity length of 8 mm and the spontaneous decay rate Γ on the $P_{1/2} \rightarrow D_{3/2}$ transition are given by $(g, \kappa, \Gamma)/2\pi = (0.92, 1.2, 1.69)$ MHz.

The first step of the protocol for generating a single-photon pulse is to apply radiation at 397 and 866 nm for 3 μs from the side of the cavity to laser-cool the ion and thus ensure its proper localization. Subsequently, the initial electronic state $S_{1/2}$ is prepared by optical pumping at 866 nm for 0.5 μs . A single photon is produced by driving a cavity-assisted Raman transition to the $D_{3/2}$ -level using a pump pulse at 397 nm with a predefined intensity profile of up to 6 μs duration. The sequence is repeated at a rate of 100 kHz.

The photon pulse is emitted from the cavity through one mirror with a transmissivity of 600 p.p.m., 100 times larger than that of the opposite mirror. The gaussian mode emanating from this mirror is focused on an avalanche photodiode (APD) after passing a series of four optical filters and a spatial filter. In this way, light from an auxiliary laser used for locking the cavity length and from other sources of stray light is reduced to a level well below the dark-count rate of our detectors. We achieve an overall single-photon detection efficiency of $(4.6 \pm 0.8)\%$. The time of each photon detection event is recorded using a multiscaler with 2-ns resolution.

With the position of the ion and hence the ion-field coupling fixed, the distribution of photon detection times exactly follows the shape of the single-photon pulse. By tailoring the intensity profile of the driving pulse, we can imprint an arbitrary temporal structure on the waveform of the photon, which is then precisely reproduced in every emission event.

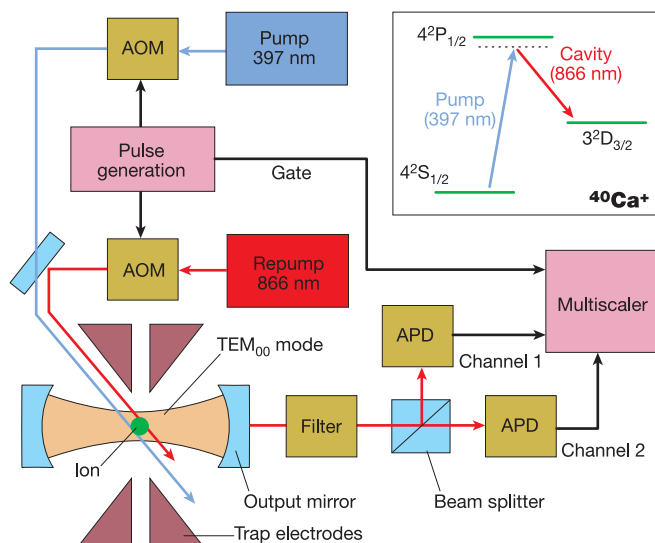


Figure 1 Experimental set-up, outlining the excitation and detection scheme for single photons. Not shown is an additional laser at 894 nm, which is resonant with a cavity mode not coupled to the ion and which is used for stabilizing the cavity length. In the path to the detectors, this locking beam is suppressed by a series of filters, providing an attenuation of 10^{11} at 894 nm. The laser intensities are controlled with acousto-optic modulators (AOMs). The inset shows the relevant levels of the $^{40}\text{Ca}^+$ ion used for single-photon production. The pump laser and cavity are red-detuned with respect to the $P_{1/2}$ -level.

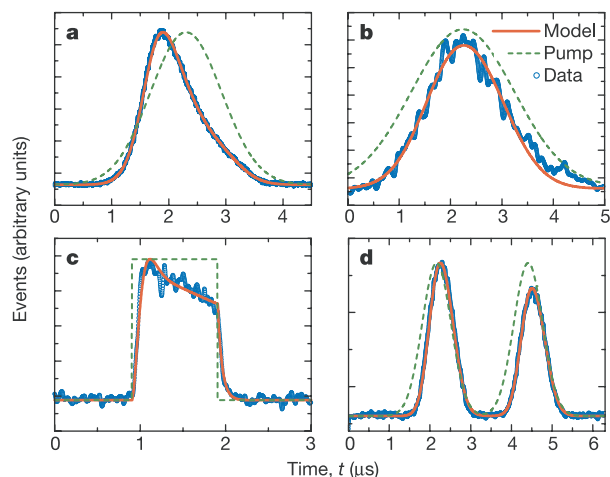


Figure 2 Single-photon pulse shapes for different pump laser profiles, indicated by the green dotted line (not to scale). **a**, Strong gaussian pump. **b**, Weak gaussian pump. **c**, Square-wave pump. **d**, Double-peaked pump. The superimposed red curves show the results of a density-matrix calculation, taking into account the full Zeeman structure of the $^{40}\text{Ca}^+$ levels, as well as the polarization of the optical fields. Plots **a** and **d**, obtained with the best statistics, demonstrate the nearly perfect agreement between the model and our data, confirming that we can deterministically control the waveform of a single photon.

The single-photon pulse shape is extracted from the time records by accumulating the probability distribution of photon arrival times relative to the pump-pulse trigger. Experimental data obtained for intense gaussian pumping (Rabi frequency $\Omega/2\pi = 10$ MHz), based on the evaluation of over 400,000 photons continuously generated from a single ion, is presented in Fig. 2a. Because the photon pulses are identical, the time distributions we have obtained constitute a measurement of the waveform of a single photon. This is confirmed by comparing the data with calculated single-photon pulse shapes, based on the coherent Raman-coupling of a calcium ion to a single cavity mode²⁴.

Precise timing and shaping of single-photon pulses is an essential requirement for realization of a bidirectional ion-photon interface in a quantum network¹³, in which the emitted pulses must be exactly symmetric in time. Figure 2b shows a symmetric gaussian output pulse, generated with a weak gaussian pump pulse ($\Omega/2\pi = 4.4$ MHz). The arbitrary control of the photon pulse shape in our set-up is limited only by the time constants associated with ion and cavity dynamics. This is apparent from the response of the system to a pump pulse with a square profile (Fig. 2c). The trailing edge of the pulse decreases exponentially at the cavity decay rate, whereas the onset of the pulse is delayed by the time required to excite the electronic transition. As an example of a more complex pulse shape, we have generated a twin-peaked single-photon pulse (Fig. 2d). The photon detection times are distributed over two well-separated maxima, reflecting the structure of the pump pulse. Spreading a single photon over two distinct time bins may be exploited as a way to encode quantum information in the time domain²⁵.

The fidelity of single-photon-state production in our system is determined from second-order temporal correlations between photons in a Hanbury-Brown-Twiss set-up. With the help of a 50% beam splitter, we randomly send the impinging photons to one of two identical APD detectors (see Fig. 1). From the arrival times, recorded separately for each detector, we determine the number of correlation events as a function of time delay, using 100-ns time bins. As explained in the Methods section, special care must be taken to eliminate detector dark-counts. The resulting correlations between photon arrival times in a range of $\pm 100 \mu\text{s}$ are shown in Fig. 3. Because photons are emitted only in response to a pump pulse, the observed correlations form a series of spikes separated by the pump period, with a width related to the photon pulse shape. The most interesting region is that around zero delay, which has contributions only if more than one photon is emitted during the

same pump pulse. The absence of a central peak in Fig. 3 proves that our system operates as a high-fidelity single-photon source.

Two factors contribute to the purity of the generated single-photon state. First, in our experiment, it is guaranteed that precisely one single ion is present in the trap at any time over many hours, in contrast to the subsecond trapping times of atoms¹⁵. Equally important, the level scheme of calcium ensures that after a photon has been deposited in the cavity, the ion occupies the metastable state $D_{3/2}$ with a one-second lifetime. This state is not coupled to the pump beam, so no additional photons are generated until the ion is actively recycled to the ground state.

To quantify the suppression of two-photon emission from our source, we evaluated the photon statistics in each pulse. The dead-time of the APDs means that each detector can reliably identify at most a single photon per excitation cycle. We therefore determined the rate of two-photon pulses from the number of events in which both detectors have recorded one photon. After correcting for dark-count events, we obtain only 2 ± 13 two-photon events during 3,000 s of data-taking, compared with $411,200 \pm 335$ recorded single photons. This is a factor of 68 better than an attenuated beam of coherent light at the same intensity. It should be noted that this value does not result from a limitation of the source (which is, in fact, expected to be entirely free of two-photon events), but is exclusively determined by the precision with which we can test the two-photon suppression in the presence of detector dark-count fluctuations. This was confirmed in a reference measurement of the dark-count statistics.

Apart from the control of the photon's waveform and the suppression of two-photon events, an important accomplishment for applications in quantum information processing is the continuous, uninterrupted operation of our source, allowing us to measure correlations on very long timescales. In Fig. 4a, we show a measurement of correlations sampled over a time of 90 min. The linear decrease of the envelope of the correlation function is a consequence of the finite duration of the measurement. No individual lines are resolved on the timescale shown, but when we zoom in at an arbitrary point on the time axis, we find the same peak structure as in Fig. 3 on either side of the origin. As an example, Fig. 4b shows correlations between photons detected 20 min apart. The operation of our system for 90 min represents the longest continuous generation of single photons from an atomic source¹⁸. With ion trapping

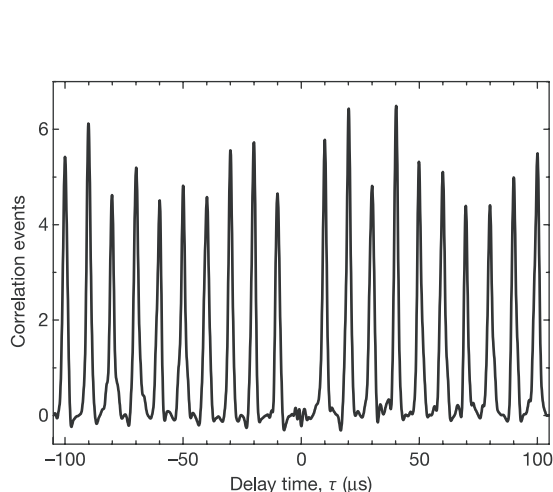


Figure 3 Cross-correlations of photon arrival times at the two detectors (counts in 100-ns time bins) around zero delay. Dark-count events have been eliminated with the help of an independent measurement (see Supplementary Fig. 1 for the data including dark-counts). The absence of a peak at $\tau = 0$ confirms that the source is emitting single photons.

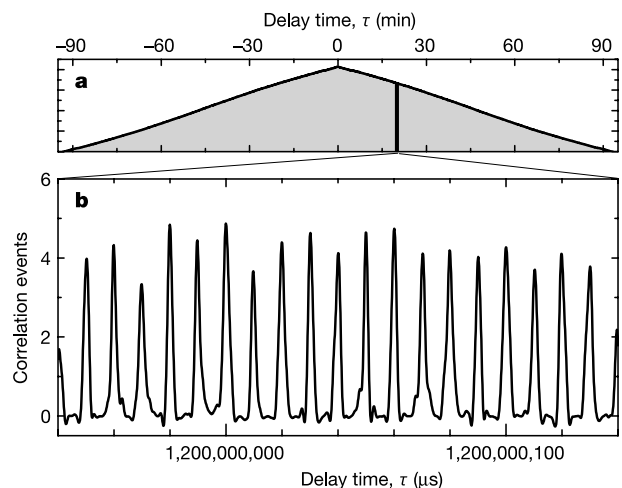


Figure 4 Long-term correlations of photon arrival times at the two detectors. **a**, Envelope of two-photon correlations recorded over the entire sampling interval, demonstrating the continuous emission of single photons in our system. **b**, Zooming in at a delay of $\tau = 20$ min reveals individual peaks as in Fig. 3 for $\tau \neq 0$.

times of many hours, single-photon emission in our system can be maintained for even longer periods.

In most of the measurements reported here, we chose the length of the pump pulses to optimize the single-photon output rate²⁴. By using slightly slower pulses and suitable detuning and intensity of the pump beam, we achieved a single-photon efficiency of $(8.0 \pm 1.3)\%$, in accordance with theoretical calculations²⁴. For the remaining 92% of the pump pulses, an infrared fluorescence photon is radiated to the side of the cavity. Note that these incoherent emission events, while reducing the efficiency of single-photon generation, do not affect the waveform of the photons emanating from the cavity. The reason is that for off-resonant Raman-coupling of the $P_{1/2}$ and $D_{3/2}$ levels, cavity emission only occurs as a result of coherent evolution.

The efficiency can be substantially increased by reducing the cavity length. When the coupling is sufficiently strong, our system constitutes a high-fidelity interface between internal quantum states of localized ions and propagating photonic states. This is an essential tool for linking quantum information processing sites in a quantum network. $\square$

Methods

The APDs we have used in evaluating the performance of our single-photon source have a large single-photon detection efficiency of 30% at 866 nm, but suffer from a non-negligible dark-count rate of approximately 50 s^{-1} . Although it is straightforward to subtract the dark-count-related background in the pulse-shape measurements of Fig. 2, the elimination of dark-count events from the correlation data requires a more elaborate procedure. If, in addition to the real photon counts $p_1(t)$ and $p_2(t)$ in the two detectors at time t , dark-count events $d_1(t)$ and $d_2(t)$ occur, the following identity holds for two-channel correlation functions:

$$\langle p_1(t+\tau)p_2(t) \rangle = \langle [p_1(t+\tau) + d_1(t+\tau)][p_2(t) + d_2(t)] \rangle - \langle d_1(t+\tau)[p_2(t) + d_2(t)] \rangle - \langle [p_1(t+\tau) + d_1(t+\tau)]d_2(t) \rangle + \langle d_1(t+\tau)d_2(t) \rangle \quad (1)$$

We determined the required correlation function $\langle p_1(t+\tau)p_2(t) \rangle$ according to equation (1) by evaluating cross-correlations between our dark-count-affected data sets ($p_1 + d_1$ or $p_2 + d_2$) and time records of dark counts only (d_1 or d_2), measured with no ion present in the trap. These reference measurements were taken immediately following the acquisition of the single-photon data. Because we averaged over a long measurement time, equation (1) is valid even though signal counts and dark counts were recorded subsequently. The four correlation measurements that were evaluated to obtain Fig. 3 are presented individually in the Supplementary Fig. 1. We took into account fluctuations of the dark-count rate by averaging over 400 different dark-count histories, unambiguously removing their contribution from the photon correlation measurements shown in Figs 3 and 4.

Received 12 May; accepted 18 August 2004; doi:10.1038/nature02961.

- Bennett, C. H. & Brassard, G. *Proc. IEEE Int. Conf. on Computers, Systems, and Signal Processing (Bangalore, India)* 175–179 (IEEE, New York, 1984).
- Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
- Diedrich, F. & Walther, H. Nonclassical radiation of a single stored ion. *Phys. Rev. Lett.* **58**, 203–206 (1987).
- Basché, T., Moerner, W. E., Orrit, M. & Talon, H. Photon antibunching in the fluorescence of a single dye molecule trapped in a solid. *Phys. Rev. Lett.* **69**, 1516–1519 (1992).
- Lounis, B. & Moerner, W. E. Single photons on demand from a single molecule at room temperature. *Nature* **407**, 491–493 (2000).
- Kurtsiefer, C., Mayer, S., Zarda, P. & Weinfurter, H. Stable solid-state source of single photons. *Phys. Rev. Lett.* **85**, 290–293 (2000).
- Brouni, R., Beveratos, A., Poizat, J.-P. & Grangier, P. Photon antibunching in the fluorescence of individual color centers in diamond. *Opt. Lett.* **25**, 1294–1296 (2000).
- Michler, P. et al. A quantum dot single-photon turnstile device. *Science* **290**, 2282–2286 (2000).
- Santori, C., Pelton, M., Solomon, G., Dale, Y. & Yamamoto, Y. Triggered single photons from a quantum dot. *Phys. Rev. Lett.* **86**, 1502–1505 (2001).
- Mureau, E. et al. Single-mode solid-state single photon source based on isolated quantum dots in pillar microcavities. *Appl. Phys. Lett.* **79**, 2865–2867 (2001).
- Yuan, Z. L. et al. Electrically driven single-photon source. *Science* **295**, 102–105 (2002).
- Santori, C., Fattal, D., Vucković, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
- Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
- Kuhn, A., Hennrich, M. & Rempe, G. Deterministic single-photon source for distributed quantum networking. *Phys. Rev. Lett.* **89**, 67901 (2002).
- McKeever, J. et al. Deterministic generation of single photons from one atom trapped in a cavity. *Science* **303**, 1992–1994 (2004).
- Brattke, S., Varcos, B. T. H. & Walther, H. Generation of photon number states on demand via cavity quantum electrodynamics. *Phys. Rev. Lett.* **86**, 3534–3537 (2001).
- Law, C. K. & Kimble, H. J. Deterministic generation of a bit-stream of single-photon pulses. *J. Mod. Opt.* **44**, 2067–2074 (1997).

- Duan, L.-M., Kuzmich, A. & Kimble, H. J. Cavity QED and quantum-information processing with “hot” trapped atoms. *Phys. Rev. A* **67**, 032305 (2003).
- Pellizzari, T., Gardiner, S. A., Cirac, J. I. & Zoller, P. Decoherence, continuous observation, and quantum computing—a cavity QED model. *Phys. Rev. Lett.* **75**, 3788–3791 (1995).
- Beige, A., Braun, D., Tregenna, B. & Knight, P. L. Quantum computing using dissipation to remain in a decoherence-free subspace. *Phys. Rev. Lett.* **85**, 1762–1765 (2000).
- Zheng, S. B. & Guo, G. C. Efficient scheme for two-atom entanglement and quantum information processing in cavity QED. *Phys. Rev. Lett.* **85**, 2392–2395 (2000).
- Guthöhrlein, G. R., Keller, M., Hayasaka, K., Lange, W. & Walther, H. A single ion as a nanoscopic probe of an optical field. *Nature* **414**, 49–51 (2001).
- Keller, M., Lange, B., Hayasaka, K., Lange, W. & Walther, H. Deterministic coupling of single ions to an optical cavity. *Appl. Phys. B* **76**, 125–128 (2003).
- Keller, M., Lange, B., Hayasaka, K., Lange, W. & Walther, H. A calcium ion in a cavity as a controlled single-photon source. *New J. Phys.* **6**, 95 (2004); available at <http://www.iop.org/EJ/abstract/1367-2630/6/1/095>.
- Brendel, J., Tittel, W., Zbinden, H. & Gisin, N. Pulsed energy-time entangled twin-photon source for quantum communication. *Phys. Rev. Lett.* **82**, 2594–2597 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We gratefully acknowledge the financial support of the European Commission through the QUEST, QUBITS and QGATES networks.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.L. (Wolfgang.Lange@mpq.mpg.de).

Crystallization of charge holes in the spin ladder of $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$

P. Abbamonte^{1,2}, G. Blumberg³, A. Ruydy^{1,4}, A. Gozar^{3,5}, P. G. Evans⁶, T. Siegrist³, L. Venema⁴, H. Eisaki⁷, E. D. Isaacs^{3,8} & G. A. Sawatzky⁹

¹National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York 11973, USA

²Department of Physics and Astronomy, SUNY Stony Brook, Stony Brook, New York 11794, USA

³Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

⁴University of Groningen, 9747 AG Groningen, The Netherlands

⁵Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

⁶Department of Materials Science & Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA

⁷Nanoelectronics Research Institute, AIST, 1-1-1 Central 2, Umezono, Tsukuba, Ibaraki, 305-8568, Japan

⁸Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA

⁹Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T-1Z1, Canada

Determining the nature of the electronic phases that compete with superconductivity in high-transition-temperature (high- T_c) superconductors is one of the deepest problems in condensed matter physics. One candidate is the ‘stripe’ phase^{1–3}, in which the charge carriers (holes) condense into rivers of charge that separate regions of antiferromagnetism. A related but lesser known system is the ‘spin ladder’, which consists of two coupled chains of magnetic ions forming an array of rungs. A doped ladder can be thought of as a high- T_c material with lower dimensionality, and has been predicted to exhibit both superconductivity^{4–6} and an insulating ‘hole crystal’^{7,8} phase in which the carriers are localized through many-body interactions. The competition between the two resembles that believed to operate between stripes and superconductivity in high- T_c materials⁹. Here we report the existence of a hole crystal in the doped spin ladder of $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ using a resonant X-ray scattering technique¹⁰. This phase exists without a detectable distortion in the

structural lattice, indicating that it arises from many-body electronic effects. Our measurements confirm theoretical predictions^{4,7,8}, and support the picture that proximity to charge ordered states is a general property of superconductivity in copper oxides.

Sr₁₄Cu₂₄O₄₁ (SCO) is a layered material consisting of two different types of copper oxide sheets—a CuO₂ ‘chain’ layer and a Cu₂O₃ ‘ladder’ layer (see ref. 11 for a picture). These two sublayers are separated by Sr atoms, and stack in an alternating fashion along the *b* crystallographic direction. The ladders and chains are parallel and run along the *c* direction, but are structurally incommensurate; that is, the ratio of their lattice parameters, $c_1/c_c = \sqrt{2}$, is not a rational number. As a result SCO is internally strained, and has a large unit cell with low-temperature lattice parameters $a = 11.47 \text{ \AA}$, $b = 13.35 \text{ \AA}$, $c = 27.3 \text{ \AA} \approx 7c_L \approx 10c_c$ (ref. 12).

SCO is an intrinsically hole-doped material with 6 hole carriers per formula unit, of which 5.2 reside in the chain layer and 0.8 in the ladder¹³. SCO has the striking property that, when alloyed with Ca and subjected to a hydrostatic pressure of 3 GPa, it superconducts with $T_c = 12 \text{ K}$ (ref. 14). Without Ca, however, it exhibits all the transport signatures of a charge density wave (CDW), including a screening mode in impedance measurements^{15,16}, a pinning mode in microwave conductivity¹⁷, a giant dielectric constant^{15,16}, and a nonlinear current–voltage (*I*–*V*) curve¹⁵, which together indicate that the carrier density is modulated in real space. These observations are typical of conventional Peierls CDW materials like NbSe₃ or K_{0.3}MoO₃ (ref. 18) in which the carrier density is modulated by a distortion in the crystal structure, driven by the electron–lattice interaction. However, a hole crystal, which is expected to compete with superconductivity in doped ladders and is driven instead by many-body interactions, would bear these same signatures. Could Sr₁₄Cu₂₄O₄₁ contain a hole crystal?

Distinguishing between the two requires determining whether the modulation is tied to a lattice distortion or occurs only among the carriers (though perhaps influencing the lattice indirectly).

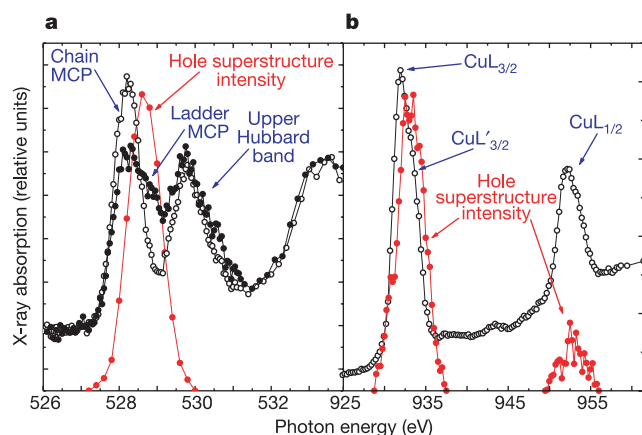


Figure 1 Energy dependence of the hole superstructure reflection compared to X-ray absorption spectra. Black symbols, absorption spectra of Sr₁₄Cu₂₄O₄₁, taken *in situ* in fluorescence-yield mode, in the vicinity of **a**, the oxygen K edge, which is a $1s \rightarrow 2p$ transition, and **b**, copper $L_{3/2,1/2}$ edges, which are $2p \rightarrow 3d$ transitions where the core hole is left with its spin either parallel ($j = 3/2$) or antiparallel ($j = 1/2$) to its orbital moment. Open circles, data taken with the photon polarization $\mathbf{E} \parallel \mathbf{a}$; filled circles, data taken with approximately $\mathbf{E} \parallel \mathbf{c}$. The data are in good agreement with ref. 13. ‘Chain MCP’ and ‘ladder MCP’ indicate the respective oxygen mobile carrier prepeaks (MCP) where scattering from the holes is enhanced. Red symbols, integrated intensity of the hole superstructure reflection as a function of incident photon energy. The reflection is visible only when the X-ray energy is tuned to the ladder MCP or the copper $L'_{3/2}$ ligand hole sideband, indicating the presence of a standing wave in the hole density in the ladder.

Recently we demonstrated that resonant X-ray scattering¹⁰ at energies near the K shell of oxygen ($1s \rightarrow 2p$ transition) is directly sensitive to hole ordering. In this method, the X-ray energy is tuned to the oxygen mobile carrier prepeak (MCP, see Fig. 1) at which scattering from the holes is selectively enhanced by $>10^3$. Here we apply this technique to search for hole ordering in SCO. This material is a particularly interesting case because it has hole carriers in both the ladder and chain layers, and the MCP is split into resolvable ladder and chain features¹³. Each provides a separate enhancement, permitting ordering in two layers to be distinguished.

Single crystals of SCO were grown by travelling solvent floating zone techniques¹⁹, cut to (0,0,1) orientation, polished with diamond film down to 0.1 μm roughness, and annealed in O₂ at 120 °C to condition the surface. Resonant soft X-ray scattering (RSXS) measurements were carried out on the X1B undulator line at the National Synchrotron Light Source with a 10-axis, ultrahigh-vacuum diffractometer. Here we will use the Miller indices *H* and *K* to

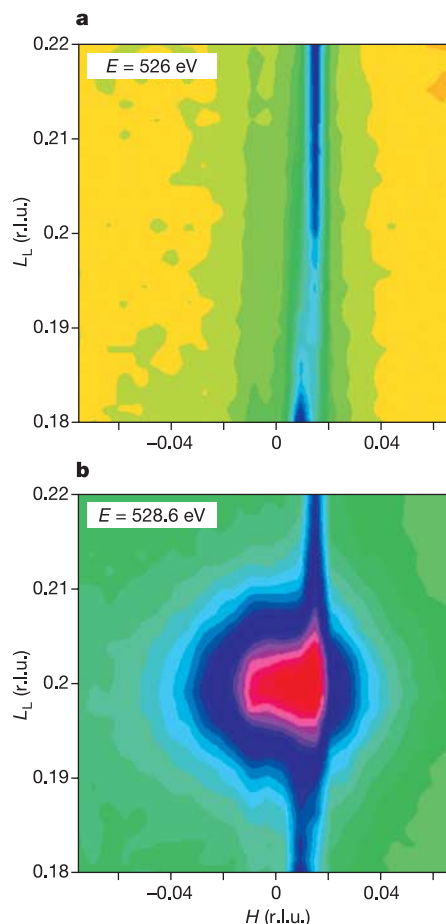


Figure 2 Appearance of the hole superstructure peak on resonance. **a**, Off-resonance ($E = 526 \text{ eV}$) reciprocal space map around $(H, K, L_L) = (0, 0, 0.2)$. The ‘rod’ at $H = 0.01$ is the specular reflectance from the surface, which is displaced from $H = 0$ because of surface miscut. The width of this rod indicates our transverse momentum resolution. **b**, Same reciprocal space region with the X-ray energy tuned to the ladder MCP ($E = 528.6 \text{ eV}$). A pronounced superlattice reflection appears at $L_L = 0.200 \pm 0.009$, indicating the presence of a commensurate standing wave in the ladder hole density with period $\lambda = 5.00c_L$. This reflection indexes to neither the $27.3 \text{ \AA}$ unit cell nor the previously reported chain dimerization reflections^{11,12,21,22}. The peak width gives longitudinal and transverse coherence lengths of $\xi_c = 255 \text{ \AA} = 65.3c_L$ and $\xi_a = 274 \text{ \AA} = 24.9a$, respectively. The hole modulation is registered across 50 neighbouring ladders, indicating significant inter-ladder coupling in this system. r.l.u., reciprocal lattice units.

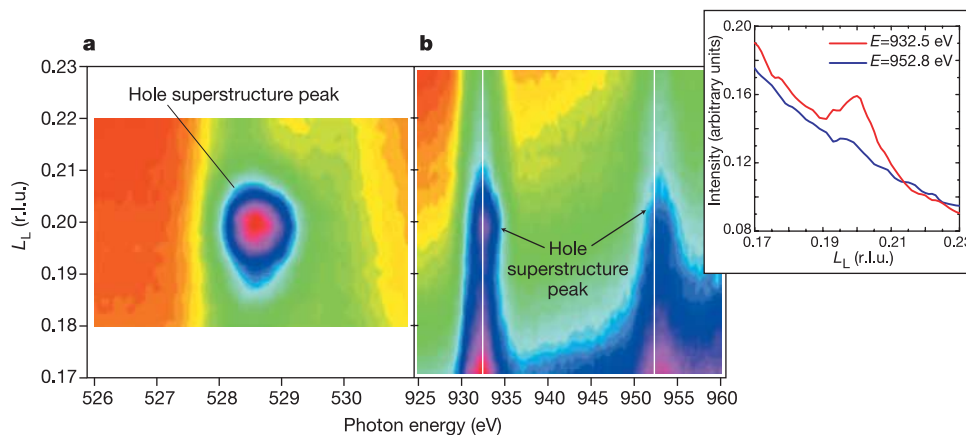


Figure 3 Energy- and L_L -dependence of the hole superstructure reflection. **a**, **b**, Peak intensity and L_L position across the oxygen K edge (**a**) and the copper $L_{3/2}$ and $L_{1/2}$ edges (**b**). The hole crystal reflection is clearly visible at both the ladder MCP (Bragg angle = 36.1°) and the copper $L_{3/2}$ edge (Bragg angle = 19.9°). In both cases it resides at $L_L = 0.200$, indicating that, although visible only at select energies, the reflection

nonetheless disperses according to Bragg's law. This establishes it as a coherent, bulk phenomenon. The background in **b** is from the specular surface reflection, which is strong at the copper L edge. Inset, L_L scans on the $L_{3/2}$ (932.5 eV) and $L_{1/2}$ (952.8 eV) peaks, corresponding to the sections indicated with white lines.

denote periodicities along the a and b directions, respectively. For the c direction we will respectively use L_c , L_L and L to denote periodicity in terms of reciprocal units of the chain, ladder, and total unit cell, that is, $L = 7L_L = 10L_c$.

Previous studies of SCO with inelastic neutron scattering²⁰ have reported evidence for a spin dimerization in the chain layer with a periodicity of $5c_c$. This has been corroborated by neutron¹², X-ray^{11,21} and electron²² diffraction, which have shown 'superlattice' Bragg reflections that index roughly as $L_c = 2n \pm 0.2$ (refs 11, 12). This phenomenon is unlikely to account for the observed CDW transport properties, however, since transport in this system is determined by the ladders²³. Moreover, in terms of the true unit cell these reflections index as $L = 20n \pm 2$ (always an integer), and so are not superlattice reflections in the true sense^{12,24}. So far, no true superlattice, with a periodicity different from the 27.3 Å unit cell, has been observed in this material.

In Fig. 2 we show reciprocal space maps around $(H, K, L_L) = (0, 0, 0.2)$ at $T = 28$ K, for X-ray energies both off and on the MCP of the ladder. Off resonance, only a specular 'rod' is visible, due to reflectance from the sample surface. If tuned to the ladder MCP, however, a pronounced superlattice reflection appears, centred at $(H, K, L_L) = (0, 0, 0.200 \pm 0.009)$, indicating the existence of a modulation along the ladder with period $5.00 \pm 0.24 c_L$. This reflection is commensurate but is truly a superlattice peak, because it occurs at $L = 1.4$ (or $L_c = 0.14$) and so does not have the periodicity of the 27.3 Å unit cell. In particular, it should not be confused with the chain dimerization reflections, which have a different periodicity.

Our central observation is the extremely unusual energy dependence of this peak. It was tracked through the oxygen K edge where it was found to be visible only for incident energies in resonance with the ladder MCP (Fig. 1)—an observation reproduced in two samples from different growth boules. The reflection is undetectable at all other energies, including the oxygen K edge jump, eliminating the possibility that it arises from a distortion in the crystal structure. In X-ray terminology, the peak responds to the anomalous scattering factors of the doped holes, and not those of the oxygen atoms, and therefore indicates a standing wave in the hole density without a (significant) lattice distortion. From its energy dependence and commensurate $5.00c_L$ periodicity, it is clear this modulation originates in the ladder substructure. The simplest interpretation is a crystallized state of holes in the ladder, which is almost degenerate with superconductivity but is stabilized under the set of parameters relevant to SCO.

To further characterize this reflection it was tracked through the L edge of copper ($2p \rightarrow 3d$ transition, Fig. 3), where it is also visible and notably still resides at $L_L = 0.2$, verifying that it disperses according to Bragg's law. Scattering at transition metal L edges is known to be sensitive to spin modulations^{25,26}, but close inspection reveals that it resonates not at the $L_{3/2}$ maximum but at the $L'_{3/2}$ shoulder, which arises from holes on the neighbouring ligands²⁷. So the modulation has no obvious magnetic character.

Finally, the X-rays were tuned to the ladder MCP and L_L scans carried out at different temperatures (Fig. 4). The hole modulation is visible below $T_c \approx 250$ K and monotonically increases with cooling. The onset is gradual but close to the $T_c = 210$ K estimated from low-frequency dielectric spectroscopy²⁸, suggesting that the hole crystal is responsible for the CDW signatures in transport. The width of the reflection is temperature-independent even near the transition, so the correlation length is limited by some mechanism

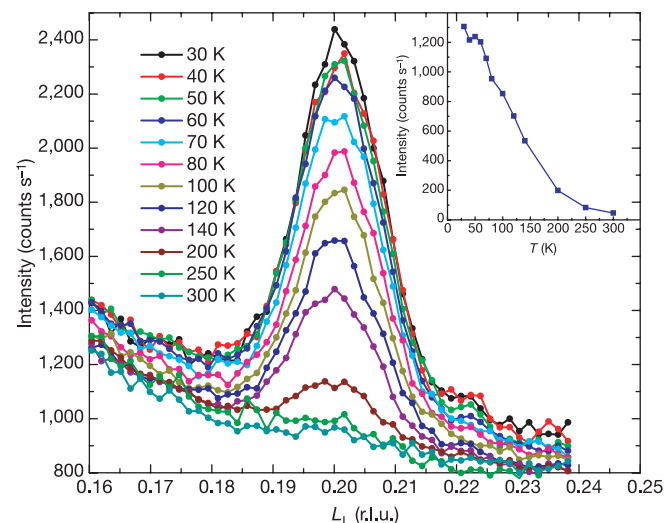


Figure 4 Temperature dependence of the hole crystal. The superlattice reflection becomes visible below $T = 250$ K. At its maximum (28 K), the peak count rate is $1,500$ photons s^{-1} on a fluorescence background of 910 photons s^{-1} . The position and width of the peak are temperature-independent. Inset, integrated intensity of the peak as a function of temperature, showing gradual, crossover behaviour, in reasonable agreement with ref. 28.

other than thermodynamics, perhaps impurities¹⁸ or intrinsic quantum fluctuations.

Our study of $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ corroborates the prediction⁴ of hole crystallization in doped ladders, and supports the picture that proximity to charge ordered states is a general property of superconductivity in copper oxides. RSXS does not permit precise determination of the form factor of the hole crystal, but no harmonic was seen at $L_L = 0.4$, suggesting a sinusoidal, delocalized modulation as discussed in ref. 7 rather than a fully localized Wigner crystal²⁹. The peak width (Fig. 2) shows that the modulation is coherent across ~ 50 neighbouring ladders, demonstrating significant inter-ladder coupling.

As a hole crystal is charged, the reader may wonder why we do not see a distortion in the lattice which might be induced electrostatically. Such a modulation must exist, but would be of the order of the amplitude of the hole modulation itself, which is probably $\sim 10^{-2}$ electrons (ref. 7). By contrast, the density modulation of a structural Peierls CDW is of the order of the atomic number, Z . So the scattering power of a hole crystal is nominally weaker by $(10^{-2}/Z)^2 \approx 10^{-6}$. Our point is not that the structural modulation is truly zero, but that electronic correlations, rather than the electron-phonon interaction, drive the transition.

A significant open question concerns the relationship between the observed wavelength of $\lambda = 5.00c_L$ and the estimated¹³ hole density in the ladder of $\delta = 0.057$ holes per copper atom. In models of hole crystallization^{7,8}, $\lambda = 1/\delta c_L$ or $2/\delta c_L$ in the strong and weak coupling regimes, respectively, which would require $\delta = 0.20$ or $\delta = 0.40$. These models neglect many residual interactions and details of the chemistry, but this relationship is resilient to such corrections. This may indicate a problem with estimates of δ , but it is worth noting that the hole crystal is commensurate with the lattice to within the measurement precision, suggesting that it is partly stabilized by elastic Umklapp processes. These can be significant for a commensurate hole crystal and perhaps strong enough to draw in extra charge from the chains. Another clue lies in the large transverse coherence length, which demonstrates significant inter-ladder interactions, and the relationship between λ and δ for a truly two-dimensional ordering pattern would not be so simple. It is therefore worth extending such models to the case of coupled ladders, or where the ladder interacts with a charge bath with which it may interchange carriers freely. □

Received 8 June; accepted 6 August 2004; doi:10.1038/nature02925.

- Zaenani, J. & Gunnarsson, O. Charged magnetic domain lines and magnetism of high- T_c oxides. *Phys. Rev. B* **40**, R7391–R7394 (1989).
- Löw, U., Emery, V. J., Fabricius, K. & Kivelson, S. A. Study of an Ising model with competing long- and short-range interactions. *Phys. Rev. Lett.* **72**, 1918–1921 (1994).
- Tranquada, J. M., Sternlieb, J. D., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper-oxide superconductors. *Nature* **375**, 561–564 (1995).
- Dagotto, E., Riera, J. & Scalapino, D. Superconductivity in ladders and coupled planes. *Phys. Rev. B* **45**, 5744–5747 (1992).
- Dagotto, E. & Rice, T. M. Surprises on the way from one- to two- dimensional quantum magnets: The ladder materials. *Science* **271**, 618–623 (1996).
- Sigrist, M., Rice, T. M. & Zhang, F. C. Superconductivity in a quasi-one-dimensional spin liquid. *Phys. Rev. B* **49**, 12058–12061 (1994).
- White, S. R., Affleck, I. & Scalapino, D. J. Friedel oscillations and charge density waves in chains and ladders. *Phys. Rev. B* **65**, 165122 (2002).
- Carr, S. T. & Tsvelik, A. M. Superconductivity and charge-density waves in a quasi-one-dimensional spin gap system. *Phys. Rev. B* **65**, 195121 (2002).
- Tranquada, J. M. et al. Coexistence of, and competition between, superconductivity and charge-stripe order in $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **78**, 338–341 (1997).
- Abbamonte, P. et al. A structural probe of the doped holes in cuprate superconductors. *Science* **297**, 581–584 (2002).
- Fukuda, T., Mizuki, J. & Matsuda, M. Periodic hole structure in a spin-chain ladder material $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. *Phys. Rev. B* **66**, 12104 (2002).
- Etrillard, J., Braden, M., Gukasov, A., Ammerahl, U. & Revcolevschi, A. Structural aspects of the spin-ladder compound $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. *Physica C* **403**, 290–296 (2004).
- Nücker, N. et al. Hole distribution in $(\text{Sr,Ca,Y,Lu})_{14}\text{Cu}_{24}\text{O}_{41}$ ladder compounds studied by x-ray absorption spectroscopy. *Phys. Rev. B* **62**, 14384–14392 (2000).
- Uehara, M. et al. Superconductivity in the ladder material $\text{Sr}_{0.4}\text{Ca}_{13.6}\text{Cu}_{24}\text{O}_{41}$. *J. Phys. Soc. Jpn* **65**, 2764–2767 (1996).
- Blumberg, G. et al. Sliding density-wave in $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ ladder compounds. *Science* **297**, 584–587 (2002).

- Gorshunov, B. et al. Charge-density wave formation in $\text{Sr}_{14-x}\text{Ca}_x\text{Cu}_{24}\text{O}_{41}$. *Phys. Rev. B* **66**, 60508(R) (2002).
- Kitano, H. et al. Microwave and millimeter wave spectroscopy in the slightly hole-doped ladders of $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. *Europhys. Lett.* **56**, 434–440 (2001).
- Grüner, G. *Density Waves in Solids* (Perseus, Cambridge, MA, 1994).
- Motomaya, N., Osafune, T., Kakeshita, T., Eisaki, H. & Uchida, S. Effect of Ca substitution and pressure on the transport and magnetic properties of $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ with doped two-leg Cu-O ladders. *Phys. Rev. B* **55**, R3386–R3389 (1997).
- Matsuda, M. et al. Magnetic excitations and structural change in the $S = 1/2$ quasi-one-dimensional magnet $\text{Sr}_{14-x}\text{Y}_x\text{Cu}_{24}\text{O}_{41}$ ($0 \leq x \leq 1$). *Phys. Rev. B* **56**, 14499–14504 (1997).
- Cox, D. E. et al. Low-temperature charge ordering in $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. *Phys. Rev. B* **57**, 10750–10754 (1998).
- Hiroi, Z., Amelinckx, S., Van Tendeloo, G. & Kobayashi, N. Microscopic origin of dimerization in the CuO_2 chains in $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. *Phys. Rev. B* **54**, 15849–15855 (1996).
- Osafune, T., Motomaya, N., Eisaki, H. & Uchida, S. Optical study of the $\text{Sr}_{14-x}\text{Ca}_x\text{Cu}_{24}\text{O}_{41}$ system: evidence for hole-doped Cu_2O_3 ladders. *Phys. Rev. Lett.* **78**, 1980–1983 (1997).
- van Smaalen, S. Comment on “Periodic hole structure in a spin-chain ladder material $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ ”. *Phys. Rev. B* **67**, 26101 (2003).
- Kao, C.-C. et al. Magnetic-resonance exchange scattering at the iron L_{II} and L_{III} edges. *Phys. Rev. Lett.* **65**, 373–376 (1990).
- Dürr, H. A. et al. Chiral magnetic domain structures in ultrathin FePd films. *Science* **284**, 2166–2168 (1999).
- Chen, C. T. et al. Out-of-plane orbital characters of intrinsic and doped holes in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **68**, 2543–2546 (1992).
- Vuletić, T. et al. Suppression of the charge-density-wave state in $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ by calcium doping. *Phys. Rev. Lett.* **90**, 257002 (2003).
- Wigner, E. On the interaction of electrons in metals. *Phys. Rev.* **46**, 1002–1011 (1934).

Acknowledgements We acknowledge J. Grazul and M. Sergent for help with sample polishing, and I. Affleck, J. B. Marston, Y.-J. Kim, P. M. Platzman, J. M. Tranquada, A. Tsvelik and T. M. Rice for discussions. This work was supported by the US Department of Energy, NWO (Dutch Science Foundation), and FOM (Netherlands Organization for Fundamental Research on Matter).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.A. (abbamonte@bnl.gov).

All-optical control of light on a silicon chip

Vilson R. Almeida, Carlos A. Barrios, Roberto R. Panepucci & Michal Lipson

School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA

Photonic circuits, in which beams of light redirect the flow of other beams of light, are a long-standing goal for developing highly integrated optical communication components^{1–3}. Furthermore, it is highly desirable to use silicon—the dominant material in the microelectronic industry—as the platform for such circuits. Photonic structures that bend, split, couple and filter light have recently been demonstrated in silicon^{4,5}, but the flow of light in these structures is predetermined and cannot be readily modulated during operation. All-optical switches and modulators have been demonstrated with III–V compound semiconductors^{6,7}, but achieving the same in silicon is challenging owing to its relatively weak nonlinear optical properties. Indeed, all-optical switching in silicon has only been achieved by using extremely high powers^{8–15} in large or non-planar structures, where the modulated light is propagating out-of-plane. Such high powers, large dimensions and non-planar geometries are inappropriate for effective on-chip integration. Here we present the experimental demonstration of fast all-optical switching on silicon using highly light-confining structures to enhance the sensitivity of light to small changes in refractive index. The transmission of the structure can be modulated by up to 94% in less than 500 ps using light pulses with energies as low as 25 pJ. These results confirm the recent theoretical prediction¹⁶ of

efficient optical switching in silicon using resonant structures.

The difficulty of modulating light using silicon structures arises from the weak dependence of the refractive index and absorption coefficient on the free-carrier concentration¹⁷. For example, for a 300- μm -long 1.55- μm Mach-Zehnder modulator based on rib waveguides with a mode-field diameter of about 5 μm , a minimum optical pump pulse energy of 2 mJ is needed to modify the real part of the refractive index by $\Delta n = -10^{-3}$ in order to achieve 100% modulation¹⁸. The absorption due to free carriers under such high powers is also small (16 dB cm^{-1} for a waveguide of rectangular cross-section, 450 nm wide and 250 nm high), which demands a straight waveguide as long as 600 μm in order to achieve a modulation depth of 90% (refs 11, 19). Liu *et al.*²⁰ have recently demonstrated a high-speed silicon optical modulator based on a MOS (metal-oxide-semiconductor) configuration; this modulator was the first high-speed optical active device on silicon—a critical stepping-stone towards an all-integrated silicon optical chip. However, owing to the weak dependence of the silicon refractive index on the free-carrier concentration, the devices in ref. 20 have relatively large lengths (of the order of millimetres).

To overcome the aforementioned limitations of silicon photonic structures, we have recently proposed the use of highly confined resonant structures for low-power light modulation by enhancing the effect of refractive index change on the transmission response¹². The results indicate that a refractive index change as small as 10^{-3} can induce a large modulation depth of 80% in a compact 20- μm -long structure. On the basis of these theoretical predictions, we present experimental results on an all-optical gate based on a silicon micrometre-size planar ring resonator, which operates with low pump-pulse energies.

The transmission of a ring resonator, coupled to a waveguide, is highly sensitive to the signal wavelength, and is greatly reduced at wavelengths at which the ring circumference corresponds to an integral number of guided wavelengths. Figure 1 shows a silicon-on-insulator (SOI) ring resonator with 10 μm diameter, patterned by electron-beam lithography and subsequently etched by inductively-coupled-plasma reactive ion etching²¹. Both the silicon waveguide and the ring resonator are channel waveguides with 450-nm-wide by 250-nm-high rectangular cross-sections. The Smart Cut SOI wafer used has a buried 3- μm -thick oxide layer. Figure 2 shows the quasi-TM transmitted spectral response of the structure in Fig. 1. The quasi-TM mode is characterized by the magnetic field being oriented predominantly along the plane of the chip. We see that on-resonance the transmitted power drops by more than 10 dB with respect to that off-resonance. The losses at off-resonance wavelengths are 3.5 dB, which include the fibre-to-waveguide coupling

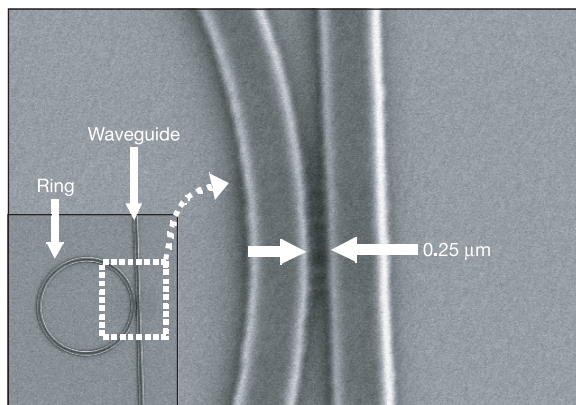


Figure 1 Scanning electron micrograph showing the top view of a ring resonator coupled to a waveguide. Inset shows the whole ring structure.

losses and the propagation losses in the 7-mm-long waveguide.

By tuning the effective index of the ring waveguide, the resonance wavelength is modified, which induces a strong modulation of the transmitted signal. Here we use 10-ps pump pulses to inject free carriers through two-photon absorption inside the ring resonator⁹, thereby tuning its effective refractive index. The probe and pump beam wavelengths are centred around two adjacent resonances of the ring resonator, $\lambda_{\text{res1}} = 1,535.6$ and $\lambda_{\text{res2}} = 1,555.5$ nm (Fig. 2), respectively. The cavity quality-factor values for these resonances are, respectively, $Q_{\text{res1}} \approx \lambda_{\text{res1}}/\Delta\lambda_{\text{FWHM1}} = 3,410$ and $Q_{\text{res2}} \approx \lambda_{\text{res2}}/\Delta\lambda_{\text{FWHM2}} = 2,290$, where $\Delta\lambda_{\text{FWHM1}} = 0.45$ nm and $\Delta\lambda_{\text{FWHM2}} = 0.68$ nm are the full-width-at-half-maximum resonance bandwidths; these correspond to cavity photon lifetimes of $\tau_{\text{cav1}} = \lambda_{\text{res1}}^2/(2\pi c\Delta\lambda_{\text{FWHM1}}) = 1.8$ ps and $\tau_{\text{cav2}} = \lambda_{\text{res2}}^2/(2\pi c\Delta\lambda_{\text{FWHM2}}) = 2.8$ ps, respectively, where c is the speed of light in vacuum²². Thus, despite the resonant nature of the structure, the temporal response of this ultra-small optical gate can theoretically be as short as a few picoseconds.

The laser source for the pump is a tunable mode-locked optical parametric oscillator, which in turn is pumped by a Ti:sapphire picosecond laser at a 78-MHz repetition rate. The optical parametric oscillator generates 1.5-ps pulses that pass through a Fabry-Perot tunable filter ($\Delta\lambda_{\text{FWHM}} = 0.37$ nm) in order to produce the pump beam, which comprises 10-ps pulses with energy of less than 25 pJ coupled to the silicon waveguide input. A tunable continuous-wave laser provides the probe signal. Both pump and probe beams are set to be linearly polarized (quasi-TM) by use of independent polarization controllers. The pump and probe beams are combined by directional couplers, and coupled into the silicon waveguide by an external tapered-lensed fibre and an on-chip fibre-to-waveguide nanotaper coupler²¹. The transmitted probe signal is coupled into a collimator, and separated from the transmitted pump light through a band-pass tunable grating filter ($\Delta\lambda_{\text{FWHM}} = 1.4$ nm). The probe signal is detected by a high-speed DC–12 GHz photodetector with a nominal fall/rise time of 30 ps. A 20-GHz digital sampling oscilloscope is used to record the probe signal.

The temporal responses of the transmitted probe signals are shown in Fig. 3 for two distinct probe wavelengths around λ_{res1} : $\lambda_{\text{probe1}} = 1,535.2$ nm (below resonance) and $\lambda_{\text{probe2}} = 1,535.6$ nm (on resonance). These probe wavelengths were tuned relative to the ring resonance in order to maximize the modulation depth by setting the transmission without pump to high and low levels,

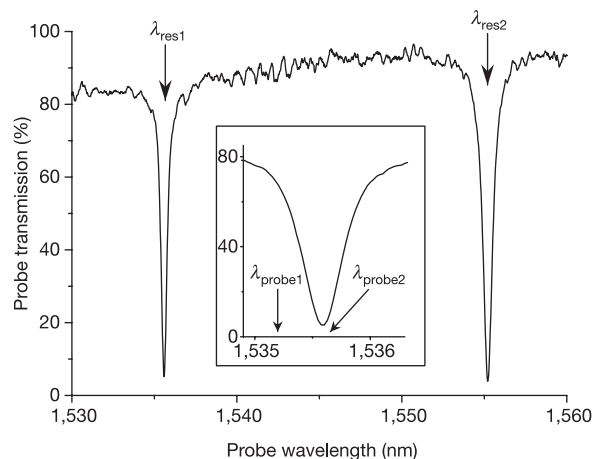


Figure 2 Quasi-TM transmission spectrum of a single-coupled ring resonator in the absence of the optical pump. Inset shows both probe wavelength settings ($\lambda_{\text{probe1}} = 1,535.2$ nm and $\lambda_{\text{probe2}} = 1,535.6$ nm) used for characterizing the dynamic response of the switch.

respectively. An important figure-of-merit for switching is the modulation depth (MD), defined as $MD = (I_{\max} - I_{\min})/I_{\max}$, where $I_{\max}$ and $I_{\min}$ are, respectively, the maximum and minimum transmitted probe optical power; we measured $MD_{\text{probe1}} = 94\%$ for λ_{probe1} and $MD_{\text{probe2}} = 91\%$ for λ_{probe2} .

By assuming an instantaneous spectral shift of the spectrum shown in Fig. 2, followed by a simple exponential decay representing the free-carrier lifetime, we obtain from the experimental data a wavelength peak shift of $\Delta\lambda = -0.36$ nm and a relaxation time of $\tau_{\text{fc}} = 450$ ps. The measured free-carrier lifetime, much shorter than that in bulk silicon, is not a fundamental limit on the speed; it is due primarily to fast recombination mechanisms on the unpassivated sidewalls of the structures. By manipulating the degree of surface passivation or by using ion implantation²³, the free-carrier lifetime could be further decreased; in this approach, using a pump time-scale much shorter than the free-carrier lifetime, the pulse energy required for operating the device remains unaltered since the switching effect occurs before the recombination process becomes significant.

The wavelength peak shift of the ring resonator corresponds to an effective index change of $\Delta n_{\text{eff}} = -4.8 \times 10^{-4}$, or equivalently to a refractive index change in the silicon core of $\Delta n_{\text{Si}} = -5.2 \times 10^{-4}$. This refractive index change is caused by a free-carrier concentration of $\Delta N = \Delta P = 1.6 \times 10^{17} \text{ cm}^{-3}$. The free-carrier concentration generated in the ring resonator is proportional to the square of the circulating peak pump power. Taking into account the volume of the ring, we estimate that the optical pulse energy absorbed inside the ring resonator in order to excite such a free-carrier concentration is only 0.15 pJ. The remaining pump power, necessary for the two-photon absorption effect, is scattered from the ring. However, this energy could be recycled by using an add/drop configuration, where an additional waveguide is added symmetrically adjacent to the ring. The losses due to the probe absorption², estimated from free-carrier concentration, are $\Delta\alpha = 6.9 \text{ cm}^{-1}$, significantly lower than the estimated scattering losses in the ring resonator of $\alpha_{\text{ring}} = 33.6 \text{ cm}^{-1}$. The relatively low absorption losses indicate that the observed modulation is due only to a refractive index change and that thermal effects can be neglected. This is of foremost importance for the application of the proposed device as an all-optical gate, enabling near 100% transmission of the data signal once the gate is open.

The device demonstrated in the present work could be used as a modulator, switch or router, with a time response as low as 100 ps.

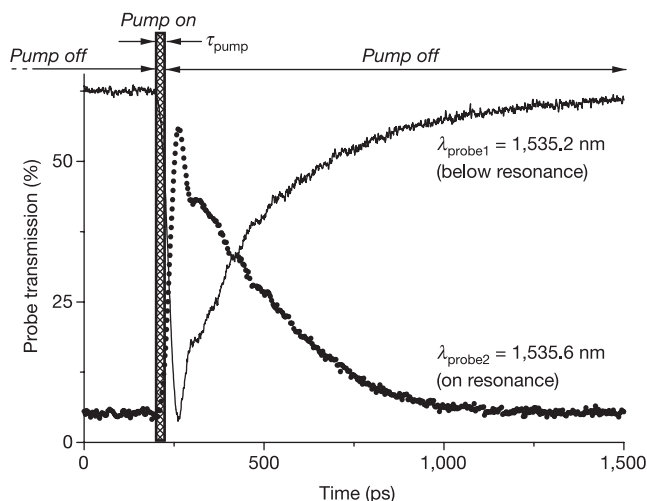


Figure 3 Temporal response of the probe signal to the pump excitation. Transmission for probe wavelengths below resonance (solid line) and on resonance (dotted line) is shown.

As a router, the device could route nanosecond-long data for reconfigurable optical interconnects²⁴. For such applications, an alternative geometry for the ring resonator, where the ring is coupled to two waveguides, could be used⁶. In this geometry, the incoming data and the control signal are coupled to the input port of the first waveguide (which contains the input port and the through port), whereas the signal output is routed to the second waveguide (the drop port). The device could switch incoming data to either the drop port or the through port, depending respectively on the presence or absence of a control pulse. For such an application, the incoming data stream would be tuned to one of the microring resonances and a control signal would be tuned to an adjacent resonance, as was done in this work.

In order to minimize the effect of the temperature variations on the device performance, strain in the silicon waveguide could be used²⁵, introduced in the fabrication process by, for example, controlling the overcladding deposition conditions²⁶. The introduced strain induces a decrease of the refractive index with temperature, which counterbalances the thermo-optic effect in silicon²⁵.

The wavelength sensitivity of the device could be decreased by minimizing the size of the ring resonators, which would result in a decrease in Q . According to our three-dimensional finite-difference time-domain simulations, ring resonators with radii as small as $0.9 \mu\text{m}$ show round-trip bending loss of less than 0.5 dB due to the high index contrast nature of the Si/SiO₂ platform; this is supported by recent experimental results²⁷. The average dissipated pump powers required for smaller resonators are similar to those required for larger resonators, in order to achieve the same modulation depth values. This is because although a larger wavelength shift is needed to obtain similar modulation depths in the smaller rings (owing to their lower Q), less pump-pulse energy is needed to obtain similar free-carrier concentrations (owing to their smaller volume).

The device described here is achieved by using the concept of strong light confinement, and is approximately seven orders of magnitude faster than available silicon optical switches²⁸. We expect that a variety of existing fabrication methods may be used to further improve the speed of the proposed device. The device shown here could form the basis for ultra-high routing bandwidth, by using architectures based on wavelength division multiplexing²⁹. □

Received 31 March; accepted 3 August 2004; doi:10.1038/nature02921.

1. Luo, C., Ioannopoulos, J. D. & Fan, S. Nonlinear photonic crystal microdevices for optical integration. *Opt. Lett.* **28**, 637–639 (2003).
2. Krauss, T. F. Planar photonic crystal waveguide devices for integrated optics. *Phys. Status Solidi A* **197**, 688–702 (2003).
3. Yablonovitch, E. Photonic crystals: semiconductors of light. *Sci. Am.* **285**, 47–55 (2001).
4. Loncar, M., Doll, T., Vuckovic, J. & Scherer, A. Design and fabrication of silicon photonic crystal optical waveguides. *J. Lightwave Technol.* **18**, 1402–1411 (2000).
5. Wada, K., Luan, H. C., Lim, D. R. C. & Kimerling, L. C. On-chip interconnection beyond semiconductor roadmap: Silicon microphotronics. *Proc. SPIE* **4870**, 437–443 (2002).
6. Ibrahim, T. A. *et al.* All-optical switching in a laterally coupled microring resonator by carrier injection. *IEEE Photon. Technol. Lett.* **15**, 36–38 (2003).
7. Van, V. *et al.* All-optical nonlinear switching in GaAs–AlGaAs microring resonators. *IEEE Photon. Technol. Lett.* **14**, 74–76 (2002).
8. Leonard, S. W., van Driel, H. M., Birner, A. & Gösele, U. All-optical ultrafast tuning of two-dimensional silicon photonic crystals via free-carrier injection. *Summaries of Papers Presented at the Quantum Electronics and Laser Science Conference. Postconference Technical Digest 159* (Optical Society of America, Washington DC, 2001).
9. Tan, H. W., van Driel, H. M., Schweizer, S. L., Wehrspohn, R. B. & Gösele, U. Tuning a 2-D silicon photonic crystal using nonlinear optics. *Conf. on Laser and Electro-Optics 2004 Vol. IFD2* (Optical Society of America, Washington DC, 2004).
10. Hache, A. & Bourgeois, M. Ultrafast all-optical switching in a silicon-based photonic crystal. *Appl. Phys. Lett.* **77**, 4089–4091 (2000).
11. Normandin, R., Houghton, D. C. & Simard-Normandin, M. All-optical, silicon based, fiber optic modulator using a near cutoff region. *Can. J. Phys.* **67**, 412–419 (1989).
12. Cocorullo, G. *et al.* Fast infrared light modulation in a-Si:H micro-devices for fiber-to-the-home applications. *J. Non-Cryst. Solids* **266–269**, 1247–1251 (2000).
13. Tsang, H. K. *et al.* Optical dispersion, two-photon absorption and self-phase modulation in silicon waveguides at 1.5 μm wavelength. *Appl. Phys. Lett.* **80**, 416–418 (2002).
14. Henari, F. Z., Morgenstern, K., Blau, W. J., Karavanskii, V. A. & Dneprovskii, V. S. Third-order optical nonlinearity and all-optical switching in porous silicon. *Appl. Phys. Lett.* **67**, 323–325 (1995).

15. Soref, R. A. & Lorenzo, J. P. Light-by-light modulation in silicon-on-insulator waveguides. *Digest of the OSA Integrated and Guided-Wave Optics Topical Meeting* 86–89 (Optical Society of America, Washington DC, 1989).
16. Barrios, C. A., Almeida, V. R. & Lipson, M. Low-power-consumption short-length and high-modulation-depth silicon electrooptic modulator. *J. Lightwave Technol.* **21**, 1089–1098 (2003).
17. Soref, R. A. & Bennett, B. R. Kramers-Kronig analysis of electro-optical switching in silicon. *Proc. SPIE* **704**, 32–37 (1987).
18. Zhao, C. Z., Li, G. Z., Liu, E. K., Gao, Y. & Liu, X. D. Silicon on insulator Mach–Zehnder waveguide interferometers operating at 1.3 μm . *Appl. Phys. Lett.* **67**, 2448–2449 (1995).
19. Stepanov, S. & Ruschin, S. Modulation of light by light in silicon-on-insulator waveguides. *Appl. Phys. Lett.* **83**, 5151–5153 (2003).
20. Liu, A. *et al.* A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor. *Nature* **427**, 615–618 (2004).
21. Almeida, V. R., Panepucci, R. R. & Lipson, M. Nanotaper for compact mode conversion. *Opt. Lett.* **28**, 1302–1304 (2003).
22. Verdeyen, J. T. *Laser Electronics*, 3rd edn 153 (Prentice Hall, Upper Saddle River, NJ, 2000).
23. Chin, A., Lee, K. Y., Lin, B. C. & Horng, S. Picosecond photoresponse of carriers in Si ion-implanted Si. *Appl. Phys. Lett.* **69**, 653–655 (1996).
24. Meindl, J. D. *et al.* Interconnect opportunities for gigascale integration. *IBM Res. Dev.* **46**, 245–263 (2002).
25. Weiss, S. M., Molinari, M. & Fauchet, P. M. Temperature stability for silicon-based photonic band-gap structures. *Appl. Phys. Lett.* **83**, 1980–1982 (2003).
26. Cheben, P., Xu, D.-X., Janz, S. & Delage, A. Scaling down photonic waveguide devices on the SOI platform. *Proc. SPIE* **5117**, 147–156 (2003).
27. Vlasov, Y. A. & McNab, S. J. Losses in single-mode silicon-on-insulator strip waveguides and bends. *Opt. Express* **12**, 1622–1631 (2004).
28. Pardo, F. *et al.* Optical MEMS devices for telecom systems. *Proc. SPIE* **5116**, 435–444 (2003).
29. Miller, D. A. B. Optical interconnects to silicon. *IEEE J. Sel. Top. Quant. Electron.* **6**, 1312–1317 (2000).

Acknowledgements We acknowledge support by the Cornell Center for Nanoscale Systems, funded by the National Science Foundation (NSF), by the Air Force Office of Scientific Research (AFOSR) and by the CS-WDM programme of the Defense Advanced Research Project Agency. V.R.A. acknowledges sponsorship support provided by the Brazilian Defence Ministry. This work was performed in part at the Cornell Nano-Scale Science & Technology Facility (CNF), a member of the National Nanotechnology Infrastructure Network (NNIN) which is supported by the NSF, its users, Cornell University and Industrial Affiliates.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.L. (lipson@ece.cornell.edu).

Unusual activity of the Sun during recent decades compared to the previous 11,000 years

S. K. Solanki¹, I. G. Usoskin², B. Kromer³, M. Schüssler¹ & J. Beer⁴

¹Max-Planck-Institut für Sonnensystemforschung (formerly the Max-Planck-Institut für Aeronomie), 37191 Katlenburg-Lindau, Germany

²Sodankylä Geophysical Observatory (Oulu unit), University of Oulu, 90014 Oulu, Finland

³Heidelberger Akademie der Wissenschaften, Institut für Umweltphysik, Neuenheimer Feld 229, 69120 Heidelberg, Germany

⁴Department of Surface Waters, EAWAG, 8600 Dübendorf, Switzerland

Direct observations of sunspot numbers are available for the past four centuries^{1,2}, but longer time series are required, for example, for the identification of a possible solar influence on climate and for testing models of the solar dynamo. Here we report a reconstruction of the sunspot number covering the past 11,400 years, based on dendrochronologically dated radiocarbon concentrations. We combine physics-based models for each of the processes connecting the radiocarbon concentration with sunspot number. According to our reconstruction, the level of solar activity during the past 70 years is exceptional, and the previous period of equally high activity occurred more than 8,000 years ago. We find that during the past 11,400 years the Sun spent only of the order of 10% of the time at a similarly high level of

magnetic activity and almost all of the earlier high-activity periods were shorter than the present episode. Although the rarity of the current episode of high average sunspot numbers may indicate that the Sun has contributed to the unusual climate change during the twentieth century, we point out that solar variability is unlikely to have been the dominant cause of the strong warming during the past three decades³.

Sunspots—strong concentrations of magnetic flux at the solar surface—are the longest-studied direct tracers of solar activity. Regular telescopic observations are available after AD 1610. In addition to the roughly 11-year solar cycle, the number of sunspots, formalized in the group sunspot number¹ (GSN), exhibits prominent fluctuations on longer timescales. Notable are an extended period in the seventeenth century called the Maunder minimum, during which practically no sunspots were present², and the period of high solar activity since about AD 1940 with average sunspot numbers above 70.

A physical approach to reconstruction of the sunspot number back in time is based on archival proxies, such as the concentration of the cosmogenic isotopes ¹⁴C in tree rings^{4–6} or ¹⁰Be in ice cores^{7,8}. This approach has recently been strengthened by the development of physics-based models describing each link in the chain of processes connecting the concentration of cosmogenic isotopes with the sunspot number^{9–12}. This advance allowed a reconstruction of the sunspot number since AD 850 based on ¹⁰Be records from Antarctica and Greenland^{13,14}. The current period of high solar activity is unique within this interval, but the covered time span is too short to judge just how unusual the current state of solar activity is.

Here we present a reconstruction of the sunspot number covering the Holocene epoch, the modern period of relatively warm climate that superseded the glacial period about 11,000 years ago. The reconstruction is based on $\Delta^{14}\text{C}$, the ¹⁴C activity in the atmosphere¹⁵ obtained from high-precision ¹⁴C analyses on decadal samples of mid-latitude tree-ring chronologies. The data set has been created in an international collaboration of dendrochronolo-

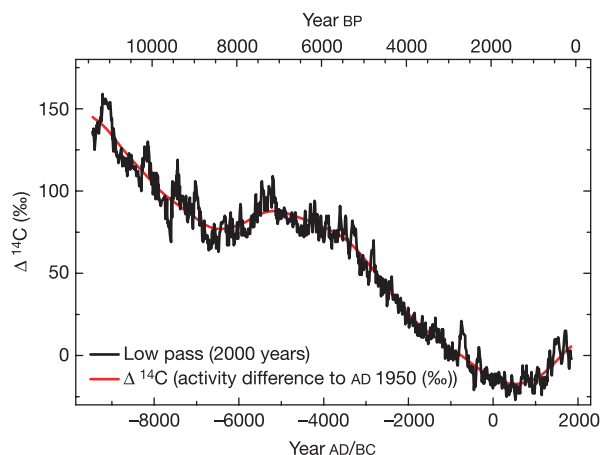


Figure 1 Atmospheric radiocarbon level $\Delta^{14}\text{C}$ (expressed as deviation, in ‰, from the AD 1950 standard level¹⁵) derived from mostly decadal samples of absolutely dated tree-ring chronologies (INTCAL98 data set)¹⁶. The $\Delta^{14}\text{C}$ measurement precision is generally 2–3‰, although in the earlier part of the time series it can reach up to 4–5‰. The INTCAL98 data for times earlier than 11,400 BP are not directly employed for the reconstruction because of larger errors and uncertainties in the carbon cycle acting at that time. See Supplementary Information for more information on the data set, initial conditions used for the reconstruction, and error estimates. The long-term decline (indicated by the red curve) is caused by a reduction in ¹⁴C production rate due mainly to an increase in the geomagnetic shielding of the cosmic ray flux. The short-term fluctuations (duration one to two centuries) reflect changes of the production rate due to solar variability. Years BC are shown negative here and in other figures.

gists and radiocarbon laboratories¹⁶. The absolutely and precisely dated original data set used for the sunspot number reconstruction is represented by the black line in Fig. 1. Starting at a level 15% higher than the reference level of AD 1950, the atmospheric ¹⁴C shows a long-term trend (indicated by the red line), which is mainly the result of changes in the intensity of the geomagnetic dipole field before and during the Holocene epoch. The fluctuations on shorter timescales predominantly result from variations of the ¹⁴C production rate due to heliomagnetic variability, which modulates the cosmic ray flux.

The atmospheric ¹⁴C level may also be affected by changes in the partition of carbon between the major reservoirs, that is, deep ocean, ocean mixed layer, biosphere and atmosphere. Variations in ocean circulation¹⁷ could influence ¹⁴C via a variable uptake of CO₂ into the ocean or by the exchange of ¹⁴C-depleted carbon from the deep ocean, but, owing to the rather small ¹⁴C gradients among the reservoirs, strong changes in these processes need to be invoked. For the Holocene, there is no evidence of considerable oceanic variability, so we can assume that the short- and mid-term fluctuations of ¹⁴C predominantly reflect solar variability. This is supported by the strong similarity of the fluctuations of ¹⁰Be in polar ice cores compared to ¹⁴C, despite their completely different geochemical history^{18–20}.

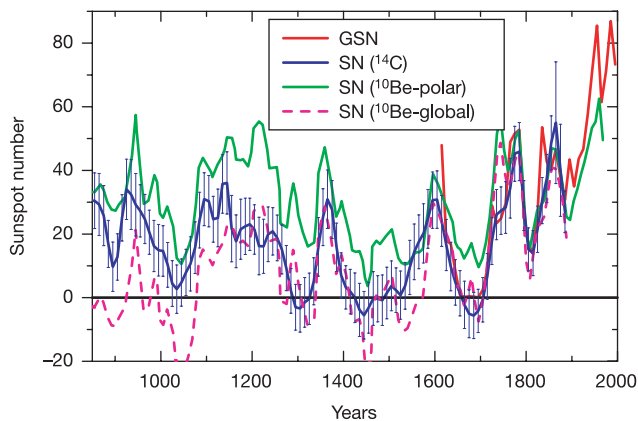


Figure 2 Comparison between directly measured sunspot number (SN) and SN reconstructed from different cosmogenic isotopes. Plotted are SN reconstructed from $\Delta^{14}\text{C}$ (blue), the 10-year averaged group sunspot number¹ (GSN, red) since 1610 and the SN reconstruction¹⁴ from ¹⁰Be under the two extreme assumptions of local (green) and global (magenta, dashed) production, respectively. The slightly negative values of the reconstructed SN during the grand minima are an artefact; they are compatible with SN = 0 within the uncertainty of these reconstructions as indicated by the error bars. $\Delta^{14}\text{C}$ is connected with the ¹⁴C production rate via a carbon cycle model²¹. The connection between the ¹⁴C production rate, R , and the cosmic ray flux is given by $R = \int_{\theta=0}^{\pi} \int_{P_c(\theta, M)}^{\infty} X(P, \Phi) Y(P) dP \sin \theta d\theta$, where θ is the colatitude relative to the geomagnetic dipole axis, and $P_c(\theta, M)$ is the local cosmic ray rigidity cutoff (which depends on θ and the virtual geomagnetic dipole moment, M)²³. $X(P, \Phi)$ is the differential cosmic ray rigidity spectrum near Earth, Φ is the modulation strength describing the average rigidity losses of cosmic rays inside the heliosphere, $Y(P)$ is the differential yield function²⁴ of ¹⁴C, and P is the rigidity of the primary cosmic rays. For studies of long-term changes of the cosmic ray flux, the parameter Φ alone adequately describes the modulation of the cosmic ray spectrum $X(P)$ ^{11,24}. The two most abundant cosmic ray species, protons and α -particles, are taken into account in the model¹³. The cosmic ray transport model relates R to Φ , which in turn depends on the Sun's open magnetic flux¹². The open flux is linked with the magnetic flux in sunspots (and thus with the SN) via the source term in a system of differential equations^{9,10}. The value of R is obtained from $\Delta^{14}\text{C}$ and M is known for the whole interval of interest^{25,26}, so that Φ can be obtained from the inversion of the equation given above. Error bars depict the 68% confidence interval for the reconstructed SN, which takes into account both random and systematic uncertainties (see Supplementary Information).

We first determine the ¹⁴C production rate in the Earth's atmosphere following Usoskin and Kromer²¹. They used two distinct methods, which take into account carbon cycle effects in different ways. Both methods give similar results when applied to the tree-ring $\Delta^{14}\text{C}$ data set described above. For the current reconstruction we use the average of the ¹⁴C production rate deduced using both techniques. In accordance with the decadal

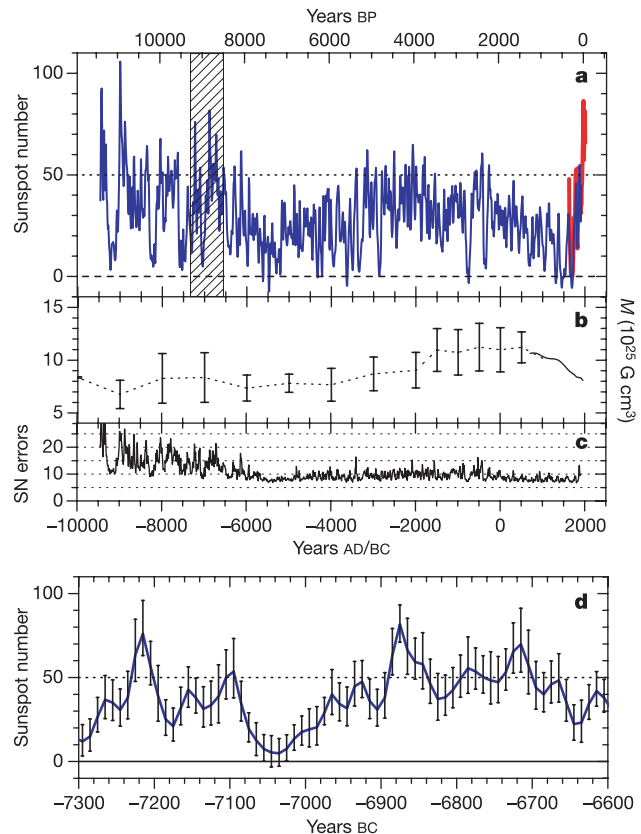


Figure 3 Reconstructed sunspot number and its uncertainty for the whole interval of time considered. **a**, 10-year averaged SN reconstructed from $\Delta^{14}\text{C}$ data since 9500 bc (blue curve) and 10-year averaged group sunspot number¹ (GSN) obtained from telescopic observations since 1610 (red curve). The horizontal dotted line marks the threshold above which we consider the Sun to be exceptionally active. It corresponds to 1.3 standard deviations above the mean. **b**, Evolution of the virtual geomagnetic dipole moment²⁶ with error bars that take into account the scatter between different palaeomagnetic reconstructions. (The error bars give the s.d. in the reconstructed virtual geomagnetic dipole moment.) The geomagnetic field data of ref. 25 are given by the dotted line. **c**, Uncertainty in the reconstructed SN. It includes errors introduced at each step of the reconstruction process. The largest sources of random errors are the uncertainty in the knowledge of the geomagnetic dipole moment and in the ¹⁴C production rate. We also consider systematic errors—for example, due to uncertainties in the ¹⁴C production rate prior to the considered period of time. A discussion of how these uncertainties are estimated is given in Supplementary Information. Clearly, the uncertainties are sufficiently small that they do not affect the presence or absence of grand minima or of episodes of high activity, except in already marginal cases. **d**, A detail from the full time series of reconstructed SN with expanded temporal scale. The chosen interval (corresponding to the shaded part of **a**) exhibits three episodes of high solar activity and a grand minimum. The error bars indicate the total uncertainty, σ , in the reconstruction. (They depict the 68% confidence interval for the reconstructed SN, which takes into account both random and systematic uncertainties (see Supplementary Information).) The two strongest maxima lie 2.1σ and 3.0σ , respectively, above the high-activity threshold of 50. Hence the probability that they are due to statistical fluctuations related to these errors is 3% and 0.2%, respectively. The probability that a whole episode of high activity (lasting, say, 50 years) is due to a statistical fluctuation is significantly smaller.

sampling of the ^{14}C data we reconstruct the 10-year averaged sunspot number. Because the $\Delta^{14}\text{C}$ data are contaminated by extensive burning of ^{14}C -free fossil fuel since the late nineteenth century²² and later by atmospheric nuclear bomb tests, we use ^{14}C data before AD 1900 only and take the historical sunspot number record for the most recent period.

From the ^{14}C production rate we obtain the sunspot number in multiple steps, each substantiated by a physics-based model. A model describing the transport and modulation of galactic cosmic rays within the heliosphere¹¹ is inverted to find the cosmic ray flux corresponding to the determined ^{14}C production rate. The transport of galactic cosmic rays in the heliosphere is affected by the Sun's open magnetic flux, that is, the fraction of the Sun's total magnetic flux that reaches out into interplanetary space¹². The open flux is linked with the sunspot number by inverting a model describing the evolution of the open magnetic flux for given sunspot number^{9,10}. All adjustable parameters entering this chain of models

were fixed using independent data prior to the current reconstruction, so that no free parameter remains when reconstructing the sunspot number from ^{14}C data (see Supplementary Table S1). This reconstruction method was previously applied to ^{10}Be data from Greenland and Antarctica. Only the first step changes when using $\Delta^{14}\text{C}$ instead of ^{10}Be data to reconstruct the sunspot number. Hence, possible errors and uncertainties in the later steps are similar to those studied in our earlier papers^{13,14}.

Applying our reconstruction method to $\Delta^{14}\text{C}$, we first determine the sunspot number since AD 850 in order to compare these values with the historical record of GSNs since 1610 and with the reconstruction on the basis of ^{10}Be data¹⁴. Figure 2 shows that the reconstructed average sunspot number from $\Delta^{14}\text{C}$ is remarkably similar to the 10-year averaged GSN series (correlation coefficient $0.925^{+0.02}_{-0.03}$ with a false alarm probability $<10^{-6}$). The difference between the reconstructed and measured sunspot number is nearly gaussian with a standard deviation of 5.8, which is smaller than the theoretical estimate of the reconstructed sunspot number uncertainty (about 8 for the last millennium, see Supplementary Information), indicating the conservative nature of the latter. Two ^{10}Be -based sunspot number reconstructions are plotted in Fig. 2, which correspond to extreme assumptions about the geographic area of ^{10}Be production relevant for its deposition in polar ice. The local polar production model (green curve) provides an upper limit to the sunspot number¹⁴, while the global production model (magenta dashed curve) gives a lower limit¹³. The sunspot number time series obtained from $\Delta^{14}\text{C}$ lies between the two ^{10}Be -based curves, and for the period after AD 1200 is closer to the ^{10}Be -based reconstruction under the assumption of global production.

Figure 3a shows the reconstruction based on the 11,400-year set of $\Delta^{14}\text{C}$ data. Clearly, the level of activity has remained variable, with episodes of particularly low numbers of sunspots (grand minima) distributed over the whole record. Episodes of high activity are also present. These are mostly concentrated in the earliest three millennia (before 6000 BC), which also exhibit a high average sunspot number (35.6 compared to 25.6 after 6000 BC). During the last eight millennia, the episode with the highest average sunspot number is the ongoing one that started about 60 years ago. The sunspot number averaged over the whole period is 28.7 with a standard deviation of 16.2. The average number of 75 since 1940 thus lies 2.85 standard deviations above this long-term average.

A major uncertainty in the reconstructed sunspot number is related to the evolution of the geomagnetic field, which is represented in Fig. 3b. A weaker geomagnetic field leads to an increased cosmic ray flux impinging on the terrestrial atmosphere and thus to a higher ^{14}C production rate, mimicking a lower value of sunspot number if not properly taken into account. The uncertainties in the reconstructed sunspot numbers are discussed in Supplementary Information and enter into Fig. 3c, where the sum of the random and systematic uncertainties affecting the reconstruction is given. For the whole time interval, the mean uncertainty of the reconstructed sunspot number is about 10.

In Fig. 3d we show a sub-interval of 700 years duration from our reconstruction, which exhibits three prominent periods of high sunspot number (average values exceeding 50) and a grand minimum. Within the 95% confidence interval, the Sun spent in total between 780 and 1,060 years in a high-activity state (sunspot number >50), which corresponds to 6.9–9.3% of the total duration of our reconstruction. The most probable values are 950 years and 8.3%, respectively. Although the rarity of the current episode of high average sunspot number may be taken as an indication that the Sun has contributed to the unusual degree of climate change during the twentieth century, we stress that solar variability is unlikely to be the prime cause of the strong warming during the last three decades³. In ref. 3, reconstructions of solar total and spectral irradiance as well as of cosmic ray flux were compared with surface temperature records

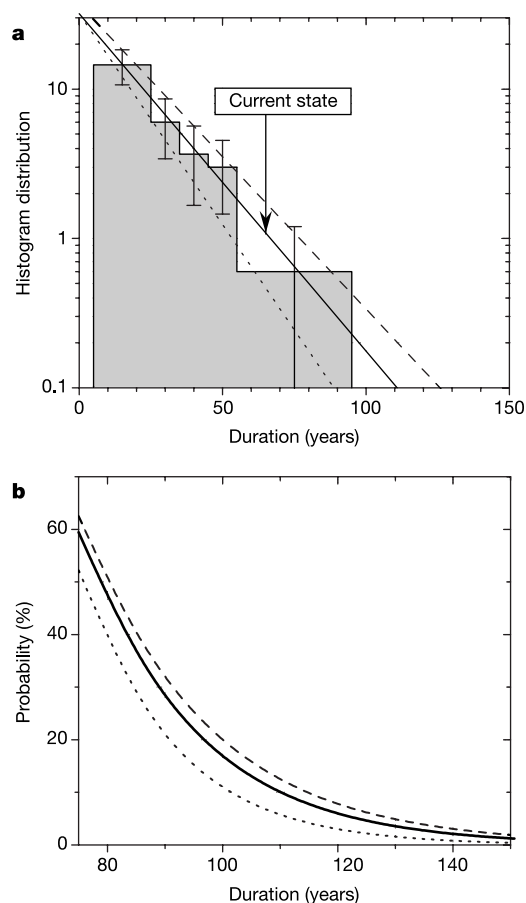


Figure 4 Distribution of the duration of episodes of high solar activity and the probability that the current episode will reach a given duration. **a**, Histogram of the distribution function of the duration of episodes of high solar activity during which the 10-year averaged SN exceeds 50. Some bins have been enlarged in order to improve the statistics; in such cases the average number per bin is given. The vertical error bars correspond to $\sqrt{N}$, where N is the number of events combined in one data point. The length of the current period of high activity is marked by the arrow. The solid line is a least-squares exponential fit to the plotted points. The dashed and dotted lines represent exponential fits to the distributions obtained from extreme SN reconstructions including the influence of random and systematic errors as given in Fig. 3 (see also the discussion in Supplementary Information). **b**, The probability of the total duration of a state of high activity (SN level exceeding 50). For the current episode, which started in AD 1940 (~65 years ago), the start of the diagram corresponds to the year AD 2015. Each curve is based upon the corresponding fit shown in **a**. The probability (for the reference curve) that the high activity continues for another 5 decades for a total duration of 115 years is only 8%.

covering approximately 150 years. It was shown that even under the extreme assumption that the Sun was responsible for all the global warming prior to 1970, at the most 30% of the strong warming since then can be of solar origin.

There are 31 periods during which the 10-year averaged sunspot number consistently exceeds a level of 50. The average length of such episodes is about 30 years, the longest being 90 years (around 9000 BC). The distribution of the durations of such episodes is given in Fig. 4a. The number of high-activity periods decreases exponentially with increasing duration. The current level of high solar activity has now already lasted close to 65 years and is marked by the arrow on the figure. This implies that not only is the current state of solar activity unusually high, but also this high level of activity has lasted unusually long. Assuming the previous episodes of high activity to be typical, we can estimate the probability with which the solar activity level will remain above a sunspot number of 50 over the next decades. The result is given in Fig. 4b, which shows that there is only a probability of $8\%^{+3\%}_{-4\%}$ that the current high-activity episode will last another 50 years (and thus reach a total duration of 115 years), while the probability that it will continue until the end of the twenty-first century is below 1%. □

Received 20 February; accepted 1 September 2004; doi:10.1038/nature02995.

- Hoyt, D. V. & Schatten, K. H. Group sunspot numbers: A new solar activity reconstruction. *Sol. Phys.* **179**, 189–219 (1998).
- Eddy, J. A. The Maunder minimum. *Science* **192**, 1189–1202 (1976).
- Solanki, S. K. & Krivova, N. Can solar variability explain global warming since 1970? *J. Geophys. Res.* **108**, doi: 10.1029/2002JA009753 (2003).
- Stuiver, M. & Braziunas, T. F. Atmospheric ^{14}C and century-scale solar oscillations. *Nature* **338**, 405–408 (1989).
- Stuiver, M. & Braziunas, T. F. Sun, ocean, climate and atmospheric $^{14}\text{CO}_2$: an evaluation of causal and spectral relationships. *Holocene* **3**, 289–305 (1993).
- Damon, P. E. & Sonett, C. P. In *The Sun in Time* (eds Sonnet, C. P., Giampapa, M. S. & Matthews, M. S.) 360–388 (Univ. Arizona, Tucson, 1991).
- Beer, J. *et al.* Use of ^{10}Be in polar ice to trace the 11-year cycle of solar activity. *Nature* **347**, 164–166 (1990).
- Beer, J. Long-term indirect indices of solar variability. *Space Sci. Rev.* **94**, 53–66 (2000).
- Solanki, S. K., Schüssler, M. & Flügge, M. Evolution of the Sun's large-scale magnetic field since the Maunder minimum. *Nature* **408**, 445–447 (2000).
- Solanki, S. K., Schüssler, M. & Flügge, M. Secular variation of the Sun's magnetic flux. *Astron. Astrophys.* **383**, 706–712 (2002).
- Usoskin, I. G., Alanko, K., Mursula, K. & Kovaltsov, G. A. Heliospheric modulation strength during the neutron monitor era. *Sol. Phys.* **207**, 389–399 (2002).
- Usoskin, I. G., Mursula, K., Solanki, S. K., Schüssler, M. & Kovaltsov, G. A. A physical reconstruction of cosmic ray intensity since 1610. *J. Geophys. Res.* **107**, doi:10.1029/2002JA009343 (2002).
- Usoskin, I. G., Mursula, K., Solanki, S. K., Schüssler, M. & Alanko, K. Reconstruction of solar activity for the last millennium using ^{10}Be data. *Astron. Astrophys.* **413**, 745–751 (2004).
- Usoskin, I. G., Solanki, S. K., Schüssler, M., Mursula, K. & Alanko, K. A millenium scale sunspot number reconstruction: evidence for an unusually active Sun since the 1940s. *Phys. Rev. Lett.* **91**, 211101 (2003).
- Stuiver, M. & Pollach, P. Discussion: reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
- Stuiver, M. *et al.* INTCAL98 Radiocarbon age calibration. *Radiocarbon* **40**, 1041–1083 (1998).
- Broecker, W. S. An unstable superconveyor. *Nature* **367**, 414–415 (1994).
- Bond, G. *et al.* Persistent solar influence on North Atlantic surface circulation during the Holocene. *Science* **294**, 2130–2136 (2001).
- Muscheler, R., Beer, J. & Kromer, B. *Solar Variability as an Input to the Earth's Environment* 305–316 (ESA SP-535, European Space Agency, Noordwijk, 2003).
- Bard, E., Raisbeck, G. M., Yiou, F. & Jouzel, J. Solar modulation of cosmogenic nuclide production over the last millennium: comparison between ^{14}C and ^{10}Be records. *Earth Planet. Sci. Lett.* **150**, 453–462 (1997).
- Usoskin, I. G. & Kromer, B. Reconstruction of the ^{14}C production rate from measured relative abundance. *Radiocarbon* (in the press).
- Suess, H. E. Radiocarbon content in modern wood. *Science* **122**, 415–417 (1955).
- Elsasser, W., Ney, E. P. & Winckler, J. R. Cosmic-ray intensity and geomagnetism. *Nature* **178**, 1226–1227 (1956).
- Castagnoli, G. & Lal, D. Solar modulation effects in terrestrial production of carbon-14. *Radiocarbon* **22**, 133–158 (1980).
- Hongre, L., Hulot, G. & Khokhlov, A. An analysis of the geomagnetic field over the past 2000 years. *Phys. Earth Planet. Inter.* **106**, 311–335 (1998).
- Yang, S., Odah, H. & Shaw, J. Variations in the geomagnetic dipole moment over the last 12000 years. *Geophys. J. Int.* **140**, 158–162 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K.S. (solanki@mps.mpg.de).

Archaeology and age of a new hominin from Flores in eastern Indonesia

M. J. Morwood¹, R. P. Soejono², R. G. Roberts³, T. Sutikna², C. S. M. Turney³, K. E. Westaway³, W. J. Rink⁴, J.-x. Zhao⁵, G. D. van den Bergh⁶, Rokus Awe Due², D. R. Hobbs¹, M. W. Moore¹, M. I. Bird⁷ & L. K. Fifield⁸

¹Archaeology and Palaeoanthropology, School of Human and Environmental Studies, University of New England, Armidale, New South Wales 2351, Australia

²Indonesian Centre for Archaeology, Jl. Raya Condet Pejaten No. 4, Jakarta 12001, Indonesia

³GeoQuEST Research Centre, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia

⁴School of Geography and Geology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

⁵Advanced Centre for Queensland University Isotope Research Excellence (ACQUIRE), University of Queensland, Brisbane, Queensland 4072, Australia

⁶Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Texel, The Netherlands

⁷School of Geography and Geosciences, University of St Andrews, St Andrews, Fife KY16 9AL, UK

⁸Research School of Physical Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia

Excavations at Liang Bua, a large limestone cave on the island of Flores in eastern Indonesia, have yielded evidence for a population of tiny hominins, sufficiently distinct anatomically to be assigned to a new species, *Homo floresiensis*¹. The finds comprise the cranial and some post-cranial remains of one individual, as well as a premolar from another individual in older deposits. Here we describe their context, implications and the remaining archaeological uncertainties. Dating by radiocarbon (^{14}C), luminescence, uranium-series and electron spin resonance (ESR) methods indicates that *H. floresiensis* existed from before 38,000 years ago (kyr) until at least 18 kyr. Associated deposits contain stone artefacts and animal remains, including Komodo dragon and an endemic, dwarfed species of *Stegodon*. *H. floresiensis* originated from an early dispersal of *Homo erectus* (including specimens referred to as *Homo ergaster* and *Homo georgicus*)¹ that reached Flores, and then survived on this island refuge until relatively recently. It overlapped significantly in time with *Homo sapiens* in the region^{2,3}, but we do not know if or how the two species interacted.

Liang Bua is a cave formed in Miocene limestone on Flores, an island in eastern Indonesia located midway between the Asian and Australian continents (Fig. 1). The cave is situated 14 km north of Ruteng and 25 km from the north coast, overlooking the Wae Racang river valley at an altitude of 500 m above sea level (08° 31' 50.4" S, 120° 26' 36.9" E). It is 30 m wide and 25 m high at the entrance, and up to 40 m deep (Fig. 2). Formed as an underground cavern by karst dissolution, the northern end was then exposed by invasion of the Wae Racang. This river now lies 200 m distant from and 30 m below Liang Bua, but five river terraces at different elevations in the valley indicate a complex process of incision over a substantial period.

Our research at Liang Bua aims to recover evidence for the history of hominin evolution, dispersal and cultural and environmental change on Flores—an island with evidence of Early Pleistocene hominin occupation by 840 kyr^{4,5}. Work involved removing backfill from four previously excavated Sectors (I, III, IV and VII) and then continuing the excavations. We have reached a maximum depth of 11 m without encountering bedrock.

Thus far, the most significant find at Liang Bua is a hominin skeleton in Sector VII, close to the east wall. Remains include a skull, mandible, pelvis and leg bones, some of which were still articulated when discovered (Fig. 3), with sufficient distinctive features to be designated a new hominin species, *Homo floresiensis*¹.

Sector VII, 2 m by 2 m in area, was excavated to red clay containing water-rolled boulders at 7.2 m depth (Fig. 4). The skeleton, together with animal remains and stone artefacts, was deposited on a gently sloping surface in dark-brown silty clay at 5.9 m depth, then covered by slope wash sediments. There was no stratigraphic or artefactual evidence for deliberate burial. The overlying layers of clay, silt and rockfall show that this slope was maintained until light-brown and grey ('white') tuffaceous silts settled in the lower, northern part of Sector VII. These tuffaceous silts were derived from volcanic eruptions and occur elsewhere in the cave, providing a useful stratigraphic marker horizon that is bracketed by ages of 13 and 11 calibrated kyr (Supplementary Table 1a) from associated charcoal, using acid-base wet oxidation, stepped-combustion (ABOX-SC) ¹⁴C (refs 6, 7 and Supplementary Information). From 4 m depth to the surface, deposits are horizontally laid and the same stratigraphic sequence extends across the cave floor, indicating a consistent pattern of sediment accumulation.

Radiocarbon and luminescence dating methods were used to infer the age of the hominin remains (Supplementary Table 1a, b), which, given their completeness and degree of articulation, must have been covered by fine sediments soon after death, when still partially fleshed. Three charcoal samples from the lowermost

excavated deposits in Sector VII were pretreated and graphitized using the ABOX-SC method, and the ¹⁴C content of the most reliable component was measured by accelerator mass spectrometry. The two samples associated with the skeleton (ANUA-27116 and ANUA-27117) yielded statistically indistinguishable calibrated ages centred on 18 kyr (68% confidence intervals: 18.7–17.9 and 18.2–17.4 cal kyr, respectively).

Luminescence dating of sediments was used to confirm the validity of these ¹⁴C ages; in particular that 'infinitely old' charcoal had not been contaminated by radiocarbon of Holocene age, resulting in the unexpectedly young ages for a hominin skeleton with so many primitive traits. Optical dating^{8,9} of potassium-rich feldspar grains, using the infrared stimulated luminescence (IRSL) emissions, yielded ages of 14 ± 2 (LBS7-40a) and 6.8 ± 0.8 (LBS7-42a) kyr for samples collected above and alongside the skeleton, respectively. Both samples exhibited significant anomalous fading (see Supplementary Information), which will cause the measured ages to be too young, but we could not reliably extend the measured fading rates to geological timescales using available fading-correction models¹⁰. Both IRSL ages, therefore, should be viewed as minimum estimates of the time since the sediments were last exposed to sunlight.

Maximum ages for sediment deposition were obtained using the light-sensitive red thermoluminescence (TL) emissions from grains of quartz^{11,12}. The TL signal is less easily bleached than the IRSL

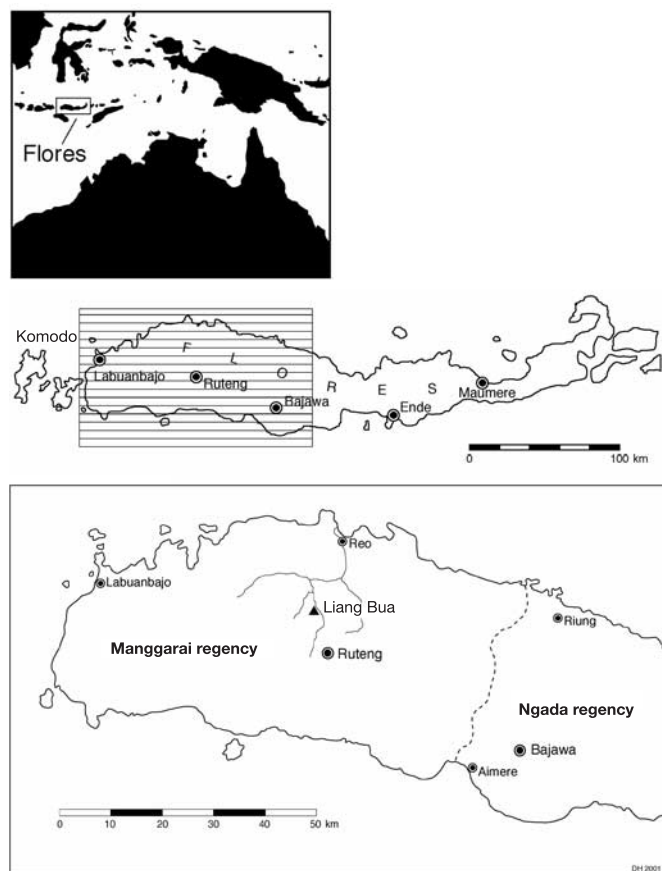


Figure 1 General location of Flores in eastern Indonesia, and Liang Bua in western Flores.

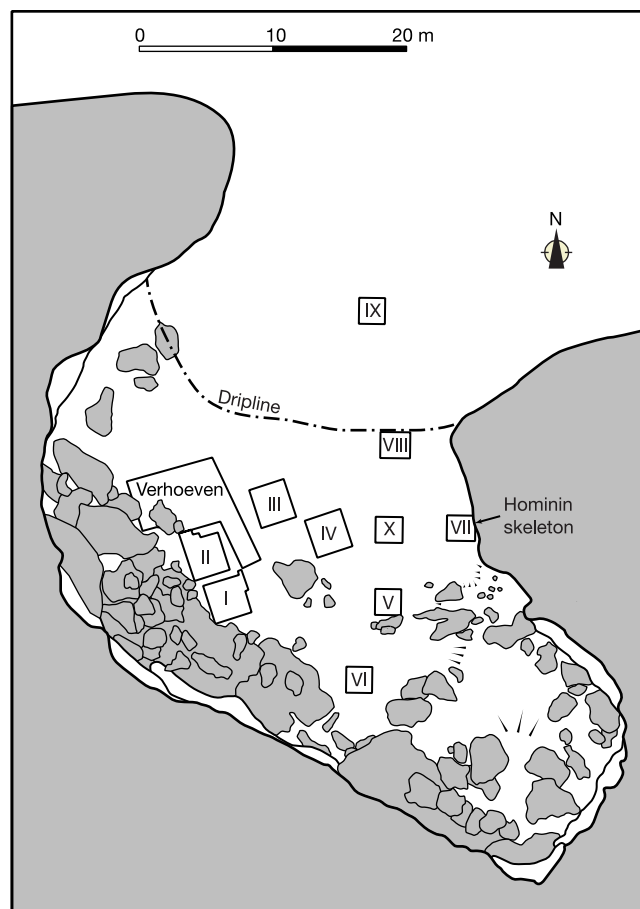


Figure 2 Plan of Liang Bua showing the locations of the excavated areas (Sectors) and the hominin skeleton (in Sector VII). Father Theodor Verhoeven carried out the first large-scale work at the site in 1965, and R. P. Soejono excavated ten Sectors between 1978 and 1989. Beginning in 2001, we extended the excavations in Sectors I, III, IV and VII.

signal, but does not suffer from anomalous fading. The TL ages for the two samples— 38 ± 8 (LBS7-40b) and 35 ± 4 (LBS7-42b) kyr—are statistically indistinguishable, supporting our contention that the body was rapidly buried soon after death. The TL and IRSR ages bracket the time of deposition of the hominin-bearing sediments to between 35 ± 4 and 14 ± 2 kyr, which is consistent with the ^{14}C ages centred on 18 kyr.

Diagnostic evidence for *H. floresiensis* is also found at Liang Bua in deposits of greater age, showing that we are not dealing with an abnormal individual but a long-standing population. At 4.3 m depth in Sector IV, deposits beneath a stratigraphic unconformity yielded a mandibular left premolar with the same distinctive morphology as premolars in the complete hominin mandible from Sector VII. Flowstone stratigraphically overlying the unconformity returned a thermal ionization mass spectrometry (TIMS) uranium-series age of 37.7 ± 0.2 kyr (sample LB-JR-6A/13–23, Supplementary Table 1c), which provides a minimum extension of the time range for *H. floresiensis*.

In addition, a juvenile *Stegodon* molar from 4.5 m depth, just below the isolated hominin premolar, yielded a coupled ESR/uranium-series age of 74^{+14}_{-12} kyr (sample LB-JR-8a, Supplementary Table 1e). Hominin remains excavated from between this dated level and 7.5 m depth, for which a maximum age of 95 ± 13 kyr for sediment deposition was obtained by TL dating (sample LBS4-32a, Supplementary Table 1b), are not yet species-diagnostic. They include, however, from a depth of 5.8 m, the radius of an adult with an estimated height of about 1 m (ref. 1) that we provisionally assign to *H. floresiensis* because of its size; the holotype lacks arms for direct comparison. If confirmed, this identification would extend the minimum antiquity of *H. floresiensis* to about 74 kyr.

Concerning the behavioural context of *H. floresiensis*, associated small faunal remains include those of fish, frog, snake, tortoise, varanids, birds, rodents and bats. Many are likely to have accumulated through natural processes, but some bones are charred, which is unlikely to have occurred naturally on a bare cave floor.

The only large animals in the Pleistocene deposits are Komodo dragon and another, even larger varanid, as well as an endemic, dwarfed species of *Stegodon*. At least 17 individuals of *Stegodon* are represented in Sector IV, and at least 9 in Sector VII. The extent of

dental wear on *Stegodon* molars also indicates that most individuals were juveniles (Age Group 1 of ref. 13), with 30% (five individuals) in Sector IV being neonates. Adults are only represented by two poorly preserved post-cranial elements and a single molar-ridge fragment. Other large mammals, such as macaque monkey, deer, pig and porcupine, first appear in the overlying Holocene deposits, which lack evidence for *H. floresiensis*. These animals were almost certainly translocated to Flores by *H. sapiens*.

Peistocene deposits in Sector VII contain relatively few stone artefacts; only 32 were found in the same level as the hominin skeleton. In Sector IV, however, dense concentrations of stone artefacts occur in the same level as *H. floresiensis*—up to 5,500 artefacts per cubic metre. Simple flakes predominate, struck bifacially from small radial cores and mainly on volcanics and chert, but there is also a more formal component found only with evidence of *Stegodon*, including points, perforators, blades and microblades that were probably hafted as barbs (Fig. 5). In all excavated Sectors, this 'big game' stone artefact technology continues from the oldest cultural deposits, dated from about 95 to 74 kyr, until the disappearance of *Stegodon* about 12 kyr, immediately below the 'white' tuffaceous silts derived from volcanic eruptions that coincide with the extinction of this species. The juxtaposition of these distinctive stone tools with *Stegodon* remains suggests that hominins at the site in the Late Pleistocene were selectively hunting juvenile *Stegodon*.

The chronologies for Sectors IV and VII show that *H. floresiensis* was at the site from before 38 kyr until at least 18 kyr—long after the 55 to 35 kyr time of arrival of *H. sapiens* in the region^{2,3,7,14–18}. None of the hominin remains found in the Pleistocene deposits, however, could be attributed to *H. sapiens*. In the absence of such evidence, we conclude that *H. floresiensis* made the associated stone artefacts.

Stone artefacts produced by much heavier percussion also occur in older deposits at Liang Bua. At the rear of the cave, for example, river-laid conglomerates contain stone artefacts, including a massive chopper. TIMS uranium-series dating of overlying flowstones indicates that these artefacts are older than 102.4 ± 0.6 kyr (sample LB-JR-10B/3–8, Supplementary Table 1c), but we do not know which hominin species manufactured them.

Further afield, the Soa Basin, which lies 50 km to the east of Liang

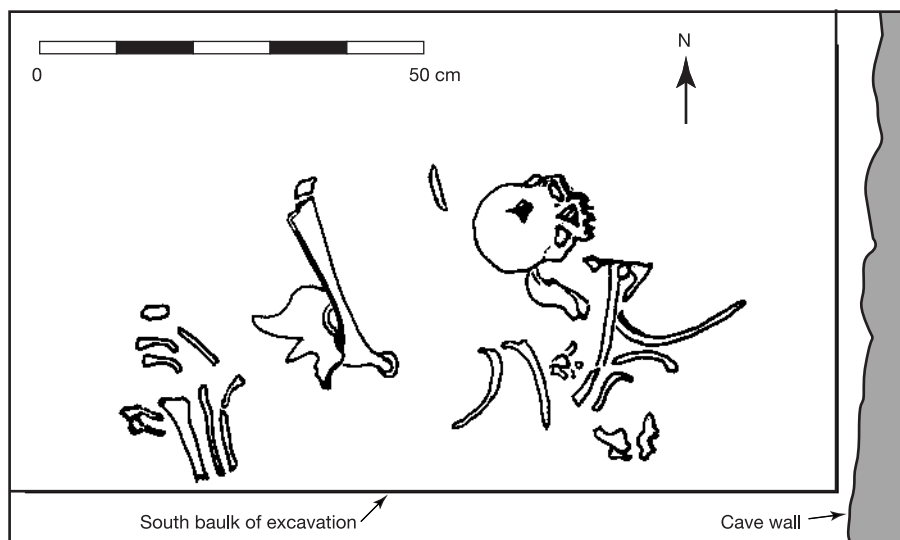


Figure 3 Plan of the hominin skeleton as found during excavation of Sector VII at Liang Bua. The relationships between skeletal elements and their proximity to the east and south baulks are shown. The right tibia and fibula were flexed beneath the corresponding femur

and patella. Additional skeletal remains, such as the arms, may lie in unexcavated deposits immediately to the south.

Bua, has sites of Early and Middle Pleistocene age, where the remains of Komodo dragon and *Stegodon* occur in association with simple, flaked stone artefacts^{4,5}. It has been assumed that *H. erectus* made these artefacts^{19–21}. The morphological traits of *H. floresiensis* at Liang Bua are consistent with *H. erectus* as an ancestral candidate, but the potential time-depth of hominin occupation of Flores means that, at this stage, we can only speculate as to which species made the Soa Basin artefacts.

Liang Bua provides evidence for distinctive hominins descended from an ancestral *H. erectus* population that survived until at least 18 kyr, overlapping significantly in time with *H. sapiens*. We interpret *H. floresiensis* as a relict lineage that reached, and was

then preserved on, a Wallacean island refuge—in the same way that Flores was a refuge for *Stegodon*, the only other large land mammal on the island during the Pleistocene. In isolation, these populations underwent protracted, endemic change; Flores was home to the smallest known species of the genera *Homo*¹ and *Stegodon*¹³.

On present evidence, the genetic and cultural isolation of Flores was only subsequently breached when *H. sapiens* appeared in eastern Asia with watercraft. How a population of tiny, small-brained hominins then survived for tens of millennia alongside *H. sapiens* remains unclear, as there is currently no evidence for the nature of their interaction; it may have involved little or no

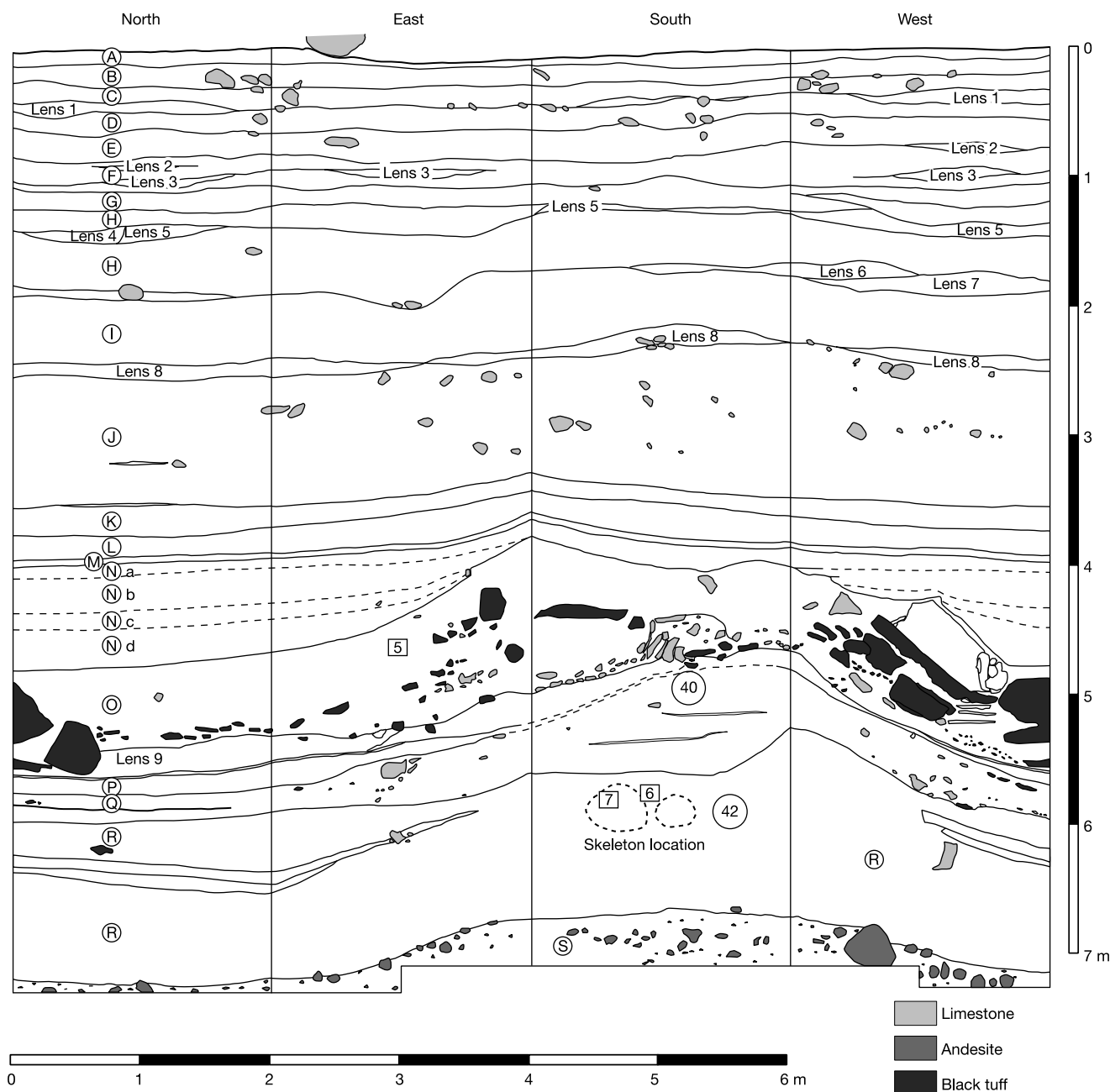


Figure 4 Stratigraphic section of the Sector VII excavation at Liang Bua, showing the location of the hominin skeleton. Layer key: A, coarse silt; B, silt; C–K, coarse silts; L, tuffaceous silt; M, clay; N (a–d), ‘white’ tuffaceous silts; O, clay and rubble; P, clay; Q, silty clay; R, sandy clay; S, clay with water-rolled volcanic boulders. The circles

enclosing the numbers 40 and 42 indicate the locations of luminescence samples LBS7-40 and LBS7-42, respectively, and the squares enclosing the numbers 5, 6 and 7 denote the locations of ¹⁴C samples ANUA-27115, ANUA-27116 and ANUA-27117, respectively.

direct contact, symbiosis, competition or predation.

The cognitive capabilities of early hominins, however, should not be underestimated, as indicated by the technology of the stone artefacts associated with *H. floresiensis* at Liang Bua. It is also significant that hominins were able to colonize Flores by the Early Pleistocene^{4,5}, whereas the required sea crossings were beyond the dispersal abilities of most other land animals, even during glacial periods of lowered sea level.

Clearly, the history of hominin occupation, evolution and cultural change on Flores, and by implication other Wallacean islands, is of much greater complexity than hitherto believed. For example, Lombok and Sumbawa are obvious stepping-stone islands for the hominin colonization of Flores from continental Asia and Java. If early hominin populations survived long-term on these islands,

they would have been subject to the same insular speciation pressures evident in *H. floresiensis*. Size reduction is a predictable evolutionary trend, but other trends will reflect island-specific adaptations, demographic changes and the impacts of catastrophic events, such as volcanic eruptions. □

Received 3 March; accepted 18 August 2004; doi:10.1038/nature02956.

1. Brown, P. *et al.* A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
2. Barker, G. *et al.* The Niah Cave Project: the second (2001) season of fieldwork. *Sarawak Mus. J.* **56** (new ser. 77), 37–119 (2001).
3. Bowler, J. M. *et al.* New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* **421**, 837–840 (2003).
4. Morwood, M. J., O'Sullivan, P. B., Aziz, F. & Raza, A. Fission-track ages of stone tools and fossils on the east Indonesian island of Flores. *Nature* **392**, 173–176 (1998).
5. Morwood, M. J. *et al.* Archaeological and palaeontological research in central Flores, east Indonesia: results of fieldwork, 1997–98. *Antiquity* **73**, 273–286 (1999).
6. Bird, M. I. *et al.* Radiocarbon dating of "old" charcoal using a wet oxidation, stepped-combustion procedure. *Radiocarbon* **41**, 127–140 (1999).
7. Turney, C. S. M. *et al.* Early human occupation at Devil's Lair, southwestern Australia 50,000 years ago. *Quaternary Research* **55**, 3–13 (2001).
8. Aitken, M. J. *An Introduction to Optical Dating* (Oxford Univ. Press, Oxford, 1998).
9. Botter-Jensen, L., McKeever, S. W. S. & Wintle, A. G. *Optically Stimulated Luminescence Dosimetry* (Elsevier Science, Amsterdam, 2003).
10. Huntley, D. J. & Lamothe, M. Ubiquity of anomalous fading in K-feldspars and the measurement and correction for it in optical dating. *Can. J. Earth Sci.* **38**, 1093–1106 (2001).
11. Franklin, A. D., Prescott, J. R. & Robertson, G. B. Comparison of blue and red TL from quartz. *Radiat. Meas.* **32**, 633–639 (2000).
12. Stokes, S. & Fattahi, M. Red emission luminescence from quartz and feldspar for dating applications: an overview. *Radiat. Meas.* **37**, 383–395 (2003).
13. van den Bergh, G. D. The Late Neogene elephantoid-bearing faunas of Indonesia and their palaeogeographic implications. A study of the terrestrial faunal succession of Sulawesi, Flores and Java, including evidence for early hominid dispersal east of Wallace's Line. *Scripta Geologica* **117**, 1–419 (1999).
14. Roberts, R. G., Jones, R. & Smith, M. A. Thermoluminescence dating of a 50,000 year-old human occupation site in northern Australia. *Nature* **345**, 153–156 (1990).
15. Roberts, R. G. *et al.* The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quaternary Sci. Rev.* **13**, 575–583 (1994).
16. Gillespie, R. Dating the first Australians. *Radiocarbon* **44**, 455–472 (2002).
17. O'Connor, S., Spriggs, M. & Veth, P. Excavation at Lene Hara Cave establishes occupation in East Timor at least 30,000–35,000 years ago. *Antiquity* **76**, 45–49 (2002).
18. O'Connell, J. F. & Allen, J. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci.* **31**, 835–853 (2004).
19. Maringer, J. & Verhoeven, Th. Die steinartefakte aus der *Stegodon*-fossilschicht von Mengeruda auf Flores, Indonesien. *Anthropos* **65**, 229–247 (1970).
20. Sondaar, P. Y. *et al.* Middle Pleistocene faunal turnover and colonisation of Flores (Indonesia) by *Homo erectus*. *C. R. Acad. Sci. Paris (Série II)* **319**, 1255–1262 (1994).
21. O'Sullivan, P. B. *et al.* Archaeological implications of the geology and chronology of the Soa Basin, Flores, Indonesia. *Geology* **29**, 607–610 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Our work is funded by a Discovery Project grant to M.J.M. from the Australian Research Council (ARC), and by grants from the University of New England (M.J.M.) and the University of Wollongong (R.G.R.). R.G.R. holds an ARC Senior Research Fellowship, and C.S.M.T. and J.-x.Z. hold ARC Queen Elizabeth II Fellowships. C.S.M.T. also acknowledges the support of the Australian Academy of Science (J. G. Russell Award), the Natural Environment Research Council and Queen's University Belfast. The 2003 excavations at Liang Bua were undertaken under Indonesian Centre for Archaeology Permit Number 1178/SB/PUS/BD/24.VI/2003. Other participants included Jatmiko, E. Wahyu Saptomo, S. Wasisto, A. Gampar, C. Lentfer, N. Polhaupessy, K. Grant, B. Walker, A. Brumm, Rikus, Deus, Leo, Ansel, Agus, Seus, Camellus, Gaba, Rius, Beni and Piet. H. Yoshida and J. Abrantes assisted with IRSL and TL analyses, J. Olley made the high-resolution gamma spectrometry measurements, D. Huntley and O. Lian provided advice on anomalous fading, and R. Bailey suggested the isothermal measurement of red TL. Wasisto, M. Roach and K. Morwood assisted with the stratigraphic sections, plans and stone artefact drawings, and P. Brown and P. Jordan commented on earlier drafts of this paper.

Author contributions M.J.M., R.P.S. and R.G.R. planned and now co-ordinate the research program funded by the ARC Discovery Project grant, which includes the Liang Bua project. T.S. directed aspects of the excavations and analyses. Ages were provided by R.G.R. and K.E.W. (luminescence); C.S.M.T., M.I.B. and L.K.F. (¹⁴C); W.J.R. (ESR); and J.-x.Z. (uranium-series). R.A.D. and G.D.v.d.B. analysed the faunal remains, and M.W.M. the stone artefacts. D.R.H. supervised the stratigraphic section drawings and other aspects of the project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.M. (mmorwood@pobox.une.edu.au) and R.G.R. (rgrob@uow.edu.au).

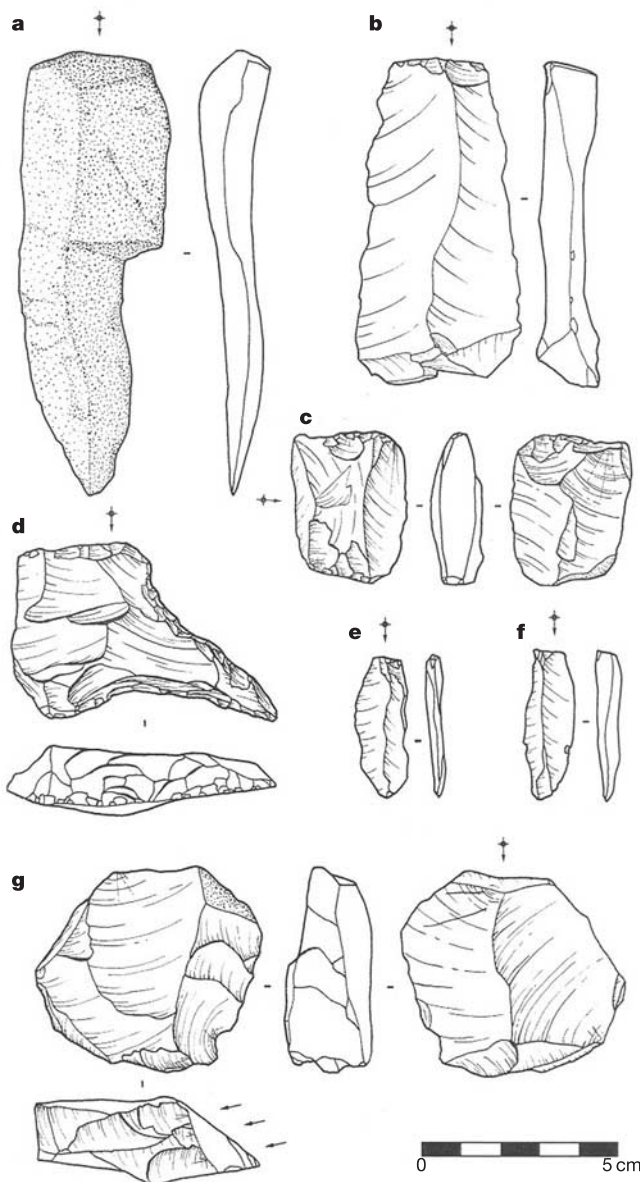


Figure 5 Range of stone artefacts associated with remains of *H. floresiensis* and *Stegodon*. **a, b**, Macroblades. **c**, Bipolar core. **d**, Perforator. **e, f**, Microblades. **g**, Burin core for producing microblades. Arrows indicate position of striking platforms, where knappers detached the flakes from cores by direct percussion using hammerstones.

Bioturbators enhance ecosystem function through complex biogeochemical interactions

Andrew M. Lohrer, Simon F. Thrush & Max M. Gibbs

National Institute of Water & Atmospheric Research, PO Box 11-115, Hillcrest, Hamilton, New Zealand

Predicting the consequences of species loss is critically important, given present threats to biological diversity such as habitat destruction, overharvesting and climate change¹. Several empirical studies have reported decreased ecosystem performance (for example, primary productivity) coincident with decreased biodiversity^{2–4}, although the relative influence of biotic effects and confounding abiotic factors has been vigorously debated^{5–7}. Whereas several investigations focused on single trophic levels (for example, grassland plants)^{8,9}, studies of whole systems have revealed multiple layers of feedbacks, hidden drivers and emergent properties^{10,11}, making the consequences of species loss more difficult to predict¹². Here we report functionally important organisms and considerable biocomplexity in a sedimentary seafloor habitat, one of Earth's most widespread ecosystems. Experimental field measurements demonstrate how the abundance of spatangoid urchins—infaunal (in seafloor sediment) grazers / deposit feeders—is positively related to primary production, as their activities change nutrient fluxes and improve conditions for production by microphytobenthos (sedimentary microbes and unicellular algae). Declines of spatangoid urchins after trawling are well documented^{13,14}, and our research linking these bioturbators to important benthic–pelagic fluxes highlights potential ramifications for productivity in coastal oceans.

Marine ecosystems are at present threatened by major anthropogenic disturbances and, although degradation of the benthos could have global consequences, the significance of species loss to benthic–pelagic rates and processes remains poorly understood^{15–17}. Benthic sediments are replete with complex biogeochemical interactions¹⁸, and ecosystem performance may depend more on the presence of key functional types than on species richness itself^{16,19}. Bioturbation by macrofauna is a function that affects sediment permeability and water content, destabilizes chemical gradients in pore water, subducts organic matter, and thus influences rates of remineralization and inorganic nutrient efflux. Because benthic habitats can supply up to half the nutrients for primary production in coastal seas, with ammonium being particularly important to nitrogen-limited marine waters, broad-scale losses of benthic bioturbators could impair marine ecosystem functioning¹⁷.

Irregular urchins of the genus *Echinocardium* are large burrowing animals common in New Zealand and similar to spatangoid urchins found throughout the world²⁰. They probably dominate bioturbation and influence sediment biogeochemistry in many locations^{21,22}, reworking surface sediment about once every 3–4 days at our study site²³, though the net effects of their activities are difficult to predict because of the number of interrelated processes involving them (Fig. 1). For example, as *Echinocardium* draws oxic water into sediments for respiration, it can increase sediment oxygen demand and by increasing bacterially and chemically mediated oxidation reactions (for example, nitrification, sulphide and pyrite oxidation)²⁴. *Echinocardium* grazes microphytobenthos and may reduce microalgal stocks, reducing oxygen production in the upper sediment column in coastal areas. However, *Echinocardium* also deposit-feeds and may reduce oxygen consuming microbes and/or their particulate organic substrates. *Echinocardium* defecates organic matter (which bacteria use in oxygen consuming reactions) and excretes

inorganic nitrogen (which microphytes use in oxygen producing reactions).

Because of their susceptibility to disturbance by bottom fishing, we performed field experiments to clarify the role of these large burrowing urchins. On each occasion, as the principal response variable, we measured oxygen production by a natural assemblage of microphytes across an experimentally controlled gradient in *Echinocardium* density. Dissolved oxygen fluxes were compared in sunlit and darkened chambers to assess net primary production and total oxygen utilization by sediments, while measurements of inorganic nutrient flux and sediment characteristics yielded insights into the mechanisms responsible for significant trends.

The natural density of *Echinocardium* at our field site was 11 ± 4 (mean per $0.25 \text{ m}^2 \pm 1 \text{ s.d.}$)²³. We deployed flux chambers to the seabed and added or removed urchins to create final densities of between 0 and 20 individuals per 0.25 m^2 chamber. Sediment texture, water content, chlorophyll *a* and organic matter content, macrofaunal density, taxonomic richness and diversity were not manipulated at our site ($\sim 25 \text{ m}^2$ extent, $\sim 6 \text{ m}$ depth). We randomized chamber positioning and collected sediment/macrofauna data from within each chamber in order to assess potentially confounding factors during statistical analysis (see Methods). These methodologies were repeated at different times of year (January and May 2003), and again after a longer-term *Echinocardium* density manipulation in January 2004.

The experimental manipulation of *Echinocardium* density was the main source of variation in our system, as *Echinocardium* dominated biomass at our site, and only two of 45 tests of

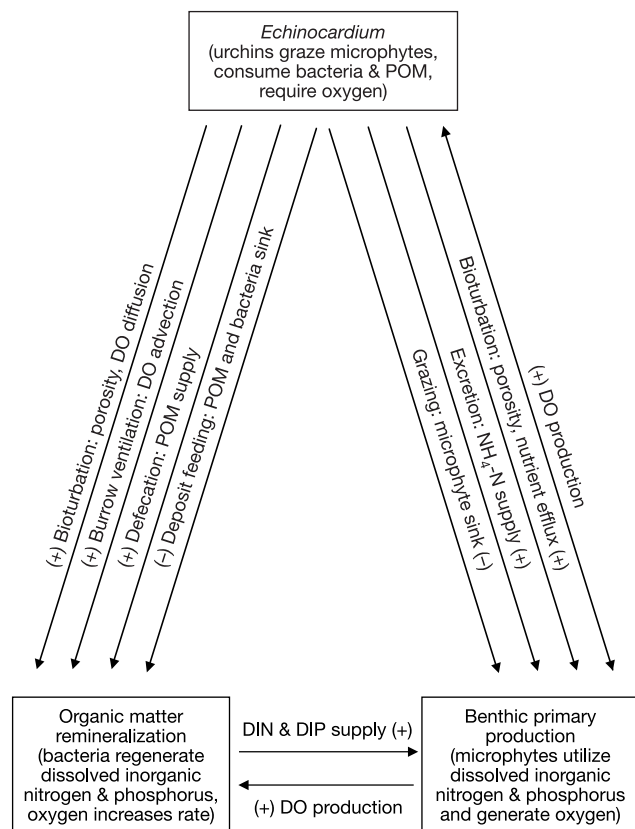


Figure 1 Diagram of interactions involving *Echinocardium*. Plus (+) indicates positive interaction (supply, facilitation). Minus (–) indicates negative interaction (sinks, inhibition). The interactions in the diagram are hypothesized (a priori); interaction strengths are not suggested. Measurements of oxygen and nutrient fluxes in sunlight and darkness are necessary to test such hypotheses. DO, dissolved oxygen; POM, particulate organic matter; DIN, dissolved inorganic nitrogen; DIP, dissolved inorganic phosphorus.

Echinocardium-parameter covariance were significant at $\alpha = 0.05$. Over time (experiment 3), *Echinocardium* enriched the sediment in organic matter ($P < 0.0001$, $r^2 = 0.82$) but had little influence on other sediment variables, including content of chlorophyll *a* ($P = 0.48$, $r^2 = 0.10$). There were no consistent effects of *Echinocardium* density on faunal variables such as macrobenthic community structure (ANOSIM²⁵, $P = 0.24$), taxonomic richness ($P = 0.65$, $r^2 = 0.02$), Shannon-Wiener diversity ($P = 0.75$, $r^2 = 0.01$), or total macrofaunal abundance ($P > 0.51$, $r^2 = 0.03$) over time. Importantly, in contrast to *Echinocardium* density, macrofaunal richness and diversity had little predictive power in multiple regression models of oxygen and nutrient flux.

Results of each experiment indicated the positive effect of *Echinocardium* on $\text{NH}_4\text{-N}$ efflux from sediments in darkened chambers (Fig. 2a, c, e). Ammonium (NH_4^+) is an animal excretory product and a nutrient used by marine microphytobenthos. Inorganic nitrogen is generally limited in marine systems, and marine plants have greater affinities for ammonium than for other forms of nitrogen such as nitrate (NO_3^-) and nitrite (NO_2^-). *Echinocardium* was a significant predictor of $\text{NH}_4\text{-N}$ efflux in experiment 2, and the correlation between *Echinocardium* density and $\text{NH}_4\text{-N}$ efflux was strongest after maintaining urchin densities for an extended period (experiment 3, $P = 0.04$, $r = 0.7$, Fig. 2e).

In two separate experiments, primary production in sunlit chambers was greatest where densities of *Echinocardium* were highest (Fig. 2d, f). This indicated the presence of complex feedbacks and interactions, considering that *Echinocardium* consumes benthic primary producers. In darkness, demand for oxygen increased with increasing numbers of *Echinocardium* (for example,

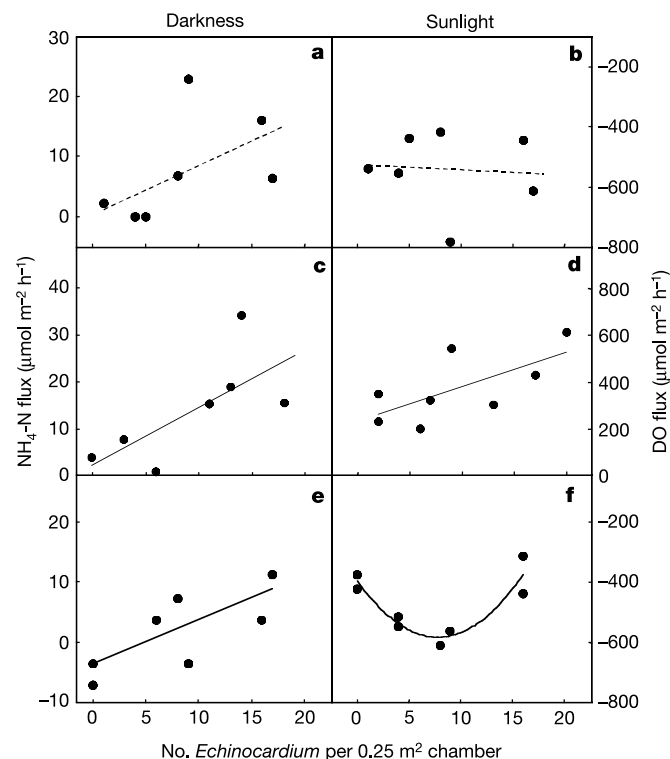


Figure 2 *Echinocardium*-mediated ammonium efflux and microphytobenthic production response. Panels **a** and **b**, **c** and **d**, and **e** and **f** are from experiments in January 2003, May 2003, and January 2004, respectively. $\text{NH}_4\text{-N}$, ammoniacal nitrogen. Positive values indicate efflux of materials out of the sediment ($\mu\text{mol m}^{-2} \text{h}^{-1}$); negative values indicate influx into sediment. *Echinocardium* was positively correlated with efflux of $\text{NH}_4\text{-N}$ in darkness in all three experiments (panels **a**, **c**, **e**; left-hand axis). Microphyte response, as indicated by DO flux in sunlit chambers, varied among experiments (panels **b**, **d**, **f**; right-hand axis). Solid lines denote significant effects of *Echinocardium* on flux.

Fig. 2a), thus the increased production of oxygen during daylight was a net result over and above the influence of animal respiration and bacterio-chemical depletion.

While similar rates of excretion probably occurred in the sunlit chambers, benthic microphytes at the sediment surface apparently intercepted and used the $\text{NH}_4\text{-N}$ as it was released from the sediment column below (for example, Fig. 3d). Uptake of $\text{NH}_4\text{-N}$ (estimated by differences in darkened and sunlit chambers) reached $20 \mu\text{mol m}^{-2} \text{h}^{-1}$ where densities of *Echinocardium* were greatest. In the absence of benthic primary production, and concordant with increased *Echinocardium* density, there was increased efflux of $\text{NH}_4\text{-N}$ from sediment to water column. Thus, when clouds, water turbidity, or water depth limit the light required for photosynthesis at the seabed, *Echinocardium* continues to provide an important ecosystem service as it subsidizes overlying waters with a readily utilizable form of inorganic nitrogen.

Although we provide evidence of increased $\text{NH}_4\text{-N}$ uptake and primary production by the microphyte community (for example,

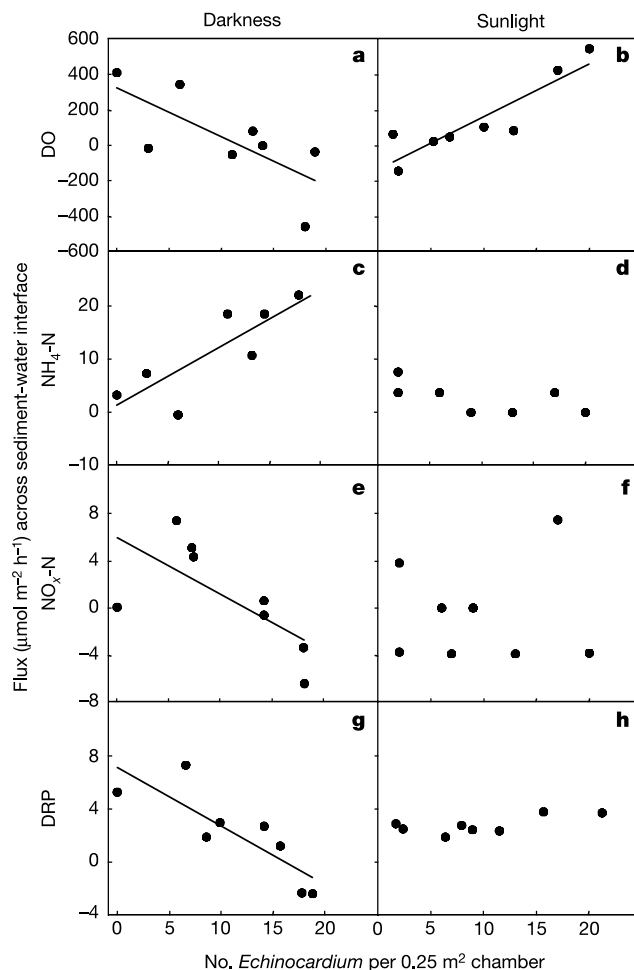


Figure 3 *Echinocardium* affects chemical fluxes. Partial regression residual plots summarize the effect of *Echinocardium* on dissolved chemical fluxes in experiment 2 after factoring out effects of other co-variables (sediment water content, organic matter content, chlorophyll *a* content, macrofaunal abundance). Dark chambers give gross trends in benthic oxygen demand and nutrient flux; sunlit chambers give net results (photosynthetic oxygen production and nutrient uptake occur in sunlight). $\text{NH}_4\text{-N}$, ammoniacal nitrogen; $\text{NO}_3\text{-N}$, nitrate-plus-nitrite nitrogen; DRP, dissolved reactive phosphorus. Residuals are centred at zero, so mean values were added back to *x*- and *y*-axis residuals and re-plotted. Presence of lines indicates significant positive or negative trends; equations from multiple regression analysis and backward selection procedures (Supplementary Table 1) give actual model fits.

Fig. 3), increased densities of *Echinocardium* did not elevate the content of chlorophyll *a* in surface sediments ($0.29 < P < 0.82$). The biogeochemical changes driven by *Echinocardium* could have shifted microphyte community composition towards species with high productivity per amount of pigment²⁶. Furthermore, grazing and bioturbation by *Echinocardium* (which would tend to remove and subduct pigment-bearing particles from surface sediment) could have negated a numerical increase in microphytes driven by increased nutrient availability/quality. For a better mechanistic understanding of these multi-layered trophic interactions (both top-down and bottom-up), detailed studies of microphyte community composition are required. Two other caveats must be considered as well. First, many pigment-containing cells (for example, benthic diatoms and flagellates) are capable of switching between autotrophy and heterotrophy depending on conditions, which complicates the relationship between chlorophyll *a* and dissolved oxygen flux. Second, there are factors besides $\text{NH}_4\text{-N}$ enhancement that could have contributed to net increases in oxygen production (for example, CO_2 enhancement, though its effects were probably minimal in this case given our study design).

In experiment 3, when *Echinocardium* had three months to condition the sediments and affect microphyte stocks, the patterns of microphyte oxygen production and chlorophyll *a* content

matched most closely. At intermediate urchin densities, the positive effects of urchin-mediated $\text{NH}_4\text{-N}$ delivery apparently did not outweigh the negative effects of urchin grazing. The presence of density dependence and nonlinear feedbacks in the provision of ecosystem services is consistent with the concept of environmental biocomplexity¹¹.

Echinocardium affected other aspects of sediment biogeochemistry as well (for example, Fig. 3e–h). In experiment 2, efflux of nitrate + nitrite nitrogen ($\text{NO}_x\text{-N}$) declined in darkened chambers as *Echinocardium* density increased (Fig. 3e). Increased consumption of dissolved oxygen with increased numbers of urchins (Fig. 3a) would tend to stimulate denitrification, an anaerobic process that converts NO_3^- and NO_2^- to N_2 gas in a series of steps. The divergent trends of decreasing $\text{NO}_x\text{-N}$ efflux and increasing $\text{NH}_4\text{-N}$ efflux (Fig. 3c, e) signalled a beneficial shift in the quality of inorganic nutrients available to the benthic microphytes. The preference of the microphytes for ammonium over nitrate was evident by comparing uptake; $\text{NH}_4\text{-N}$ uptake increased with increasing primary productivity, whereas $\text{NO}_x\text{-N}$ uptake declined. The net efflux of $\text{NO}_x\text{-N}$ in sunlit chambers (despite the potential for uptake by microphytes) indicated an additional source of dissolved inorganic nitrogen for the water column.

In the presence of oxygen, phosphate will chemically adsorb onto clay particles, binding with the positively charged edges of the particles and also replacing silicates inside the clay matrix²⁴. Because phosphate easily adsorbs to cations and forms insoluble precipitates in oxic conditions, the ventilation of sediments with oxic water by *Echinocardium* (to meet its respiratory needs) was predicted to reduce concentrations of dissolved reactive phosphorus (DRP). Increased *Echinocardium* densities decreased DRP efflux in darkened chambers during experiment 2 (Fig. 3g), though the influence of *Echinocardium* in light chambers was not significant (Fig. 3h).

An anticipated goal of future biocomplexity research is a greater understanding of the role of living organisms in ecosystem performance and in local, regional and global geochemical cycles¹¹. Sand and mud habitats account for approximately 70% of the marine sea floor, and the importance of the soft-sediment benthos is increasingly recognized for its contribution to the productivity of overlying waters and to global elemental budgets^{15,17,27}. Spatangoid urchins are found in marine soft-sediments worldwide, and can clearly affect the appearance (Fig. 4) and functioning of such systems when abundant. Easily destroyed by anthropogenic disturbance^{13,14,28}, their losses may have lasting consequences for important benthic-pelagic processes and indicators of ecosystem performance such as primary and secondary production. Unlike temperate grassland systems, where high biodiversity imparts a degree of 'insurance' against the loss of ecosystem functioning¹, reductions in density of a single key species in our system (of 45 macrofaunal taxa collected) had broad biogeochemical implications, including reduced primary production. □

Methods

Location and dates

Experiments were conducted in a soft-bottom habitat in Otarawao Bay, New Zealand ($36^\circ 30.75' \text{ S}$, $174^\circ 43.57' \text{ E}$), in January 2003, May 2003 and January 2004 (experiments 1, 2 and 3, respectively). Sediment texture was dominated by fine + very fine sand (70–75%; 63–250 μm) with a silt + clay fraction of 15–25% (<63 μm). Bottom water averaged 21 and 17 $^\circ\text{C}$ during the experiments in January and May, respectively.

Repetition and modification

All three experiments were of identical design (see below), though actions were taken before experiment 3 to examine longer-term effects of *Echinocardium*. In early November 2003, eight plots were established at the study site. The thin aluminium borders of these plots (5 m circumference) were used to fence in *Echinocardium* at fixed densities. A significant gradient in urchin density (0–21 urchins per 0.25 m^2 quadrat) was maintained for 3 months, and the benthic flux chambers of experiment 3 (January 2004) were positioned in these *Echinocardium*-conditioned sediments.

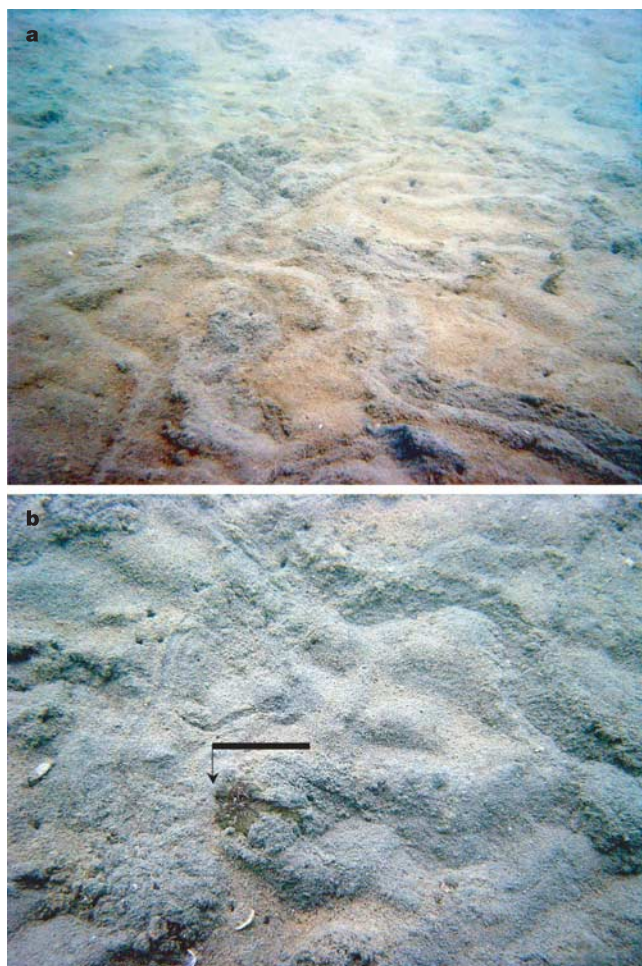


Figure 4 Photographs of sea floor, 6 m depth, Otarawao Bay, North Island, New Zealand. Panel **a** shows how the bioturbation activities of *Echinocardium* clearly influence sediment topography, with many other concomitant effects. The top 5 cm of sediment is reworked about once every three days during summer at this site²³. Arrow in **b** indicates anterior end of an adult *Echinocardium*; scale bar in this close-up shows approximate body length (3 cm).

Experimental design

Aluminium chamber bases were pressed 10 cm into the sediment to isolate square patches of sea floor 50 × 50 cm. Sixteen chamber bases were established in close proximity at a depth of 6 m. *Echinocardium* were either added or removed from the bases to create experimental treatments with 0, 4, 8 or 16 urchins per chamber. The following morning, chamber lids (8 clear and 8 opaque) with non-directional water stirrers were fitted, enclosing about 25 l of bottom water. The light–dark treatment was interspersed equally across the urchin density treatment in a randomized block design. Chamber water was sampled near midday at 1.5–2 h intervals (exact times noted in each instance), providing the raw data for flux calculations 24 h after *Echinocardium* density manipulations. Bottom water external to the chambers was also sampled at each interval, and light and dark bottles were established just above the seabed at time = 0 on the day of sampling. Dissolved oxygen was measured within minutes of collection (YSI model 5730 BOD bottle probe) and water samples were filtered immediately thereafter (1.1 µm pore size Whatman GF/C glass fibre filter). All samples were kept in darkness, and stored frozen until analysis.

Water chemistry

Analysis for ammoniacal nitrogen (NH₄-N), nitrate-plus-nitrite nitrogen (NO_x-N) and dissolved reactive phosphorus (DRP) used standard methods for sea water²⁹ on an AlphKem series 500 air-segmented continuous flow auto-analyser; detection limits <0.1 µmol l⁻¹ for N and P.

Sediments and macrofauna

To characterize features of the sediment column that could potentially affect pore water chemistry and flux chamber measurements, several types of sediment samples were collected from each chamber at the end of each experiment. Two surface sediment samples (~30 g scrapes to 2 cm depth) were collected, one for chlorophyll *a* analysis and one for organic matter content and sediment texture analysis. Chlorophyll *a* was extracted from sediment by boiling in 95% ethanol and analysed spectrophotometrically³⁰. Organic matter content was assessed by % sediment mass lost following combustion (% loss on ignition, LOI). Sediment texture was assessed by standard sieve and pipette techniques after removal of organic matter (digestion in 9% hydrogen peroxide). Water content in the upper 5 cm of sediment was determined from one 2.4 cm diameter core per chamber from the difference in sample wet weight and dry weight. Macrofauna were collected with one 10 cm diameter, 13 cm deep core taken from the centre of each chamber base. Macrofauna were sieved on a 500 µm mesh sieve and preserved in 70% isopropyl alcohol + rose bengal for later sorting and identification. When all sediment samples had been collected, the entire area enclosed by each chamber base was excavated to a depth of approximately 5 cm in order to quantify all *Echinocardium* present at time = end.

Data analysis

Each experiment was analysed separately. Sediment and faunal variables were used as predictors of dissolved chemical fluxes in multiple regression models. Variables were eliminated by backward selection unless significant at $\alpha = 0.15$. Collinearity among predictor variables was avoided by examining variance inflation factors and condition indices. Homogeneity of variance was evaluated by plotting residuals versus predicted values, and normality was assessed via normal probability plots and Shapiro-Wilk tests on residuals, though no data transformations were required.

Received 27 May; accepted 20 September 2004; doi:10.1038/nature03042.

Published online 6 October 2004.

1. Tilman, D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **80**, 1455–1474 (1999).
2. Naeem, S., Thompson, L., Lawler, S., Lawton, J. & Woodfin, R. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–736 (1994).
3. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
4. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **295**, 639–640 (1999).
5. Naeem, S., Loreau, M. & Inchausti, P. in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 3–11 (Oxford Univ. Press, Oxford, 2002).
6. Huston, M. & McBride, A. in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 47–60 (Oxford Univ. Press, Oxford, 2002).
7. Tilman, D., Knops, J., Wedin, D. & Reich, P. in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 21–35 (Oxford Univ. Press, Oxford, 2002).
8. Hector, A., Loreau, M. & Schmid, B. in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 36–46 (Oxford Univ. Press, Oxford, 2002).
9. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
10. Polis, G. Stability is woven by complex webs. *Nature* **395**, 744–745 (1998).
11. Michener, W. *et al.* Defining and unraveling biocomplexity. *Bioscience* **51**, 1018–1023 (2001).
12. Raffaelli, D. *et al.* in *Biodiversity and Ecosystem Functioning* (eds Loreau, M., Naeem, S. & Inchausti, P.) 147–154 (Oxford Univ. Press, Oxford, 2002).
13. Thrush, S. F. *et al.* Disturbance of the marine benthic habitat by commercial fishing: Impacts at the scale of the fishery. *Ecol. Appl.* **8**, 866–879 (1998).
14. Jennings, S., Pinnegar, J. K., Polunin, N. V. C. & Warr, K. J. Impacts of trawling disturbance on the trophic structure of benthic invertebrate communities. *Mar. Ecol. Prog. Ser.* **213**, 127–142 (2001).
15. Snelgrove, P. *et al.* The importance of marine sediment biodiversity in ecosystem processes. *Ambio* **26**, 578–583 (1997).
16. Bolam, S., Fernandes, T. & Huxham, M. Diversity, biomass, and ecosystem processes in the marine benthos. *Ecol. Monogr.* **72**, 599–615 (2002).
17. Thrush, S. F. & Dayton, P. K. Disturbance to marine benthic habitats by trawling and dredging — implications for marine biodiversity. *Annu. Rev. Ecol. Syst.* **33**, 449–473 (2002).

18. Herman, P., Middleburg, J., VandeKoppel, J. & Heip, C. Ecology of estuarine benthos. *Adv. Ecol. Res.* **29**, 165–231 (1999).
19. Thrush, S., Hewitt, J., Gibbs, M., Lundquist, C. & Norkko, A. Functional role of large organisms in intertidal communities: community effects and ecosystem function. *Ecology* (submitted).
20. Ghiold, J. Species distributions of irregular echinoids. *Biol. Oceanogr.* **6**, 79–162 (1989).
21. Long, B. G., Poiner, I. R. & Wassenberg, T. J. Distribution, biomass and community structure of megabenthos of the Gulf of Carpentaria, Australia. *Mar. Ecol. Prog. Ser.* **129**, 127–139 (1995).
22. Thompson, B., Tsukada, D. & Laughlin, J. Megabenthic assemblages of coastal shelves, slopes, and basins off southern California. *Bull. S. Calif. Acad. Sci.* **92**, 25–42 (1993).
23. Lohrer, A. M., Thrush, S. F., Hunt, L., Hancock, N. & Lundquist, C. J. Rapid reworking of subtidal sediments by burrowing spatangoid urchins. *J. Exp. Mar. Biol. Ecol.* (submitted).
24. Valiela, I. *Marine Ecological Processes* (Springer, New York, 1995).
25. Clark, K. R. & Gorley, R. N. *Primer version 5* (Primer-E, Plymouth, UK, 2001).
26. Falkowski, P. & Kiefer, D. A. Chlorophyll *a* fluorescence in phytoplankton: relationship to photosynthesis and biomass. *J. Plankt. Res.* **7**, 715–731 (1985).
27. Gray, J. S. Marine biodiversity: patterns, threats and conservation needs. *Biodivers. Conserv.* **6**, 153–175 (1997).
28. Daan, R. & Mulder, M. On the short-term and long-term impact of drilling activities in the Dutch sector of the North Sea. *ICES J. Mar. Sci.* **53**, 1036–1044 (1996).
29. Grasshoff, K., Ehrhardt, M. & Kremling, F. *Methods of Seawater Analysis* (Verlag Chemie, Weinheim, 1983).
30. Sartory, D. P. *Spectrophotometric Analysis of Chlorophyll *a* in Freshwater Phytoplankton* (Technical Report TR 115, Hydrological Research Institute, Department of Environmental Affairs, Pretoria, South Africa, 1982).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the New Zealand Foundation for Research Science and Technology and NIWA-NSOF for financial support. R. Budd, L. Chiaroni, V. Cummings, G. Funnell, J. Halliday, N. Hancock, J. Hewitt, C. Lundquist, A. Norkko and K. Vopel provided assistance with diving, fieldwork, laboratory analyses, and comments on earlier drafts.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.L. (d.lohrer@niwa.co.nz).

Effect of extrinsic mortality on the evolution of senescence in guppies

David N. Reznick¹, Michael J. Bryant^{1,2}, Derek Roff¹, Cameron K. Ghalambor³ & Dionna E. Ghalambor³

¹Department of Biology, University of California, Riverside, California 92521, USA

²School of Critical Studies, California Institute of the Arts, 24700 McBean Parkway, Valencia, California 91355, USA

³Department of Biology, Colorado State University, Fort Collins, Colorado 80523-8037, USA

Classical theories^{1,2} for the evolution of senescence predict that organisms that experience low mortality rates attributable to external factors, such as disease or predation, will evolve a later onset of senescence. Here we use patterns of senescence in guppies derived from natural populations that differ in mortality risk to evaluate the generality of these predictions. We have previously found that populations experiencing higher mortality rates evolve earlier maturity and invest more in reproduction, as predicted by evolutionary theory³. We report here that these same populations do not have an earlier onset of senescence with respect to either mortality or reproduction but do with respect to swimming performance, which assesses neuromuscular function. This mosaic pattern of senescence challenges the generality of the association between decreased extrinsic mortality and delayed senescence and invites consideration of more derived theories for the evolution of senescence.

Medawar's¹ 'mutation accumulation' theory predicts that populations with high mortality rates should accumulate deleterious mutations that reduce fitness late in life. This age specificity occurs

because few individuals survive long enough to experience the purifying effects of natural selection that would otherwise remove late-acting mutations from the gene pool. Williams² 'antagonistic pleiotropy' theory predicts that high mortality rates will select for earlier maturity and a higher rate of investment in reproduction early in life, which incurs a cost in the form of reduced investment in maintenance and reproduction late in life. More derived theories for the evolution of senescence generate variations on these predictions. For example, age classes may vary in their susceptibility to extrinsic mortality⁴ or there may be an interaction between senescence and vulnerability to mortality factors, such that senescence selectively increases the probability of death in one subset of the population more than others^{5,6}. Increases in extrinsic mortality rate may be accompanied by decreases in population density and increases in resource availability to survivors⁵. When such complexities are included, increased extrinsic mortality may cause the evolution of earlier senescence, later senescence, or no change in patterns of senescence, depending on these additional factors.

There is a striking discrepancy between the diversity of theory on the evolution of senescence and its treatment in the literature. Empirical evaluations of the evolution of senescence focus almost exclusively on the classical theory^{7–16}, as do recent reviews^{17,18}.

Here we report a comparative study of senescence based on the quantification of complete life histories of 240 individually reared guppies (*Poecilia reticulata*) derived from high and low mortality rate environments in Trinidad. High mortality rate sites are streams where predators co-occur with guppies. These predators are often excluded from the upper reaches of streams by waterfalls, giving rise to low mortality rate sites. Predators increase the mortality rates of all size/age classes of guppies, and guppies cannot outgrow susceptibility to predation. The probability of surviving for six months in the low predation sites is 20–30 times greater than in high predation sites^{19,20}.

We compared senescence among paired high and low predation

localities from two drainages, the Oropuche and Yarra. Previous research suggests that guppies have adapted to high and low predation environments independently in each drainage²¹. Our comparisons were made between second-generation, laboratory reared offspring derived from wild-caught females. Two generations of lab rearing removes confounding environmental effects. The offspring are the product of a breeding design that equally represents 25 wild-caught adult females, so they are free of adaptation to the laboratory or inbreeding²² (see Methods). Fish were reared individually, beginning at an age of 25–30 days and ending at death. Data collected included the age at first reproduction, age and size at all subsequent reproductive events, number of offspring and age at death. In order to compare populations in an environment as free as possible of extrinsic mortality that comes as close as possible to replicating nature, we matched water quality, temperature and photoperiod to prevailing conditions in nature, following the methods of our earlier work²².

We reared the fish at two levels of food availability to match average differences among high and low predation localities in nature²². Low predation localities tend to have lower levels of food availability²³, reflected in the lower growth rates and smaller asymptotic body sizes of guppies from those sites. Our high and low food treatments yielded asymptotic body sizes that approximated what we observed in high and low predation environments (data not shown). Thus, the complete factorial experiment was a comparison of guppies from high and low mortality rate environments in two drainages, for a total of four localities, crossed with two levels of food availability.

Whereas theory is clear in making predictions about how senescence should evolve, it is less clear about how one should quantify senescence. We have taken literally the definition of senescence as any age-specific decline in variables associated with individual fitness, specifically mortality, reproduction and physiological performance.

High predation guppies matured at a significantly earlier age, reproduced more frequently and produced more offspring in each litter²⁰, which is consistent with our earlier studies^{3,22}. Guppies provide no postnatal care for their young, so the age at last reproduction is equivalent to dying with respect to the determination of an individual's fitness. Because analyses of age at last reproduction and age at death yield qualitatively similar results, we

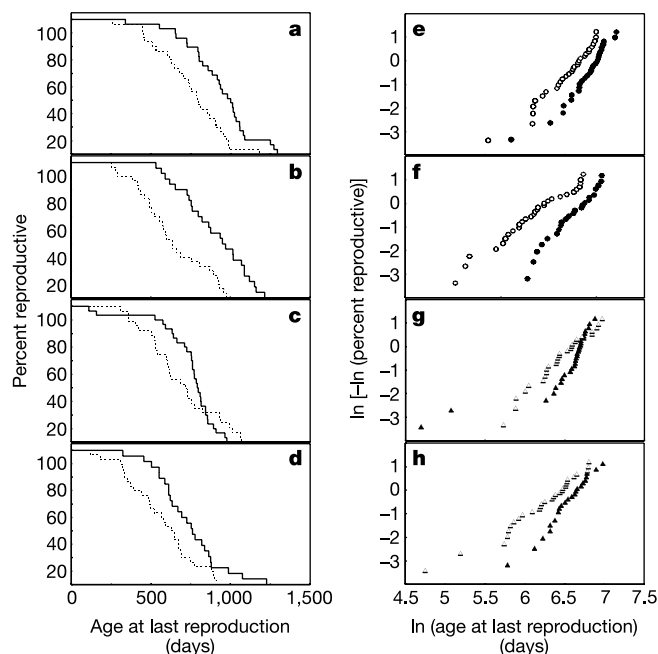


Figure 1 Age at last reproduction. **a–d**, The proportion of the population that is still reproducing at a given age (*y*-axis) plotted against the age at last reproduction (*x*-axis). High predation (solid line), low predation (stippled line). **e–h**, The natural log of negative natural log of the percent of fish that are still reproducing (*y*-axis) plotted against the age at last reproduction (natural log transformed, *x*-axis). High predation (filled symbols), low predation (open symbols). Oropuche, high food (**a**, **e**). Oropuche, low food (**b**, **f**). Yarra, high food (**c**, **g**). Yarra, low food (**d**, **h**).

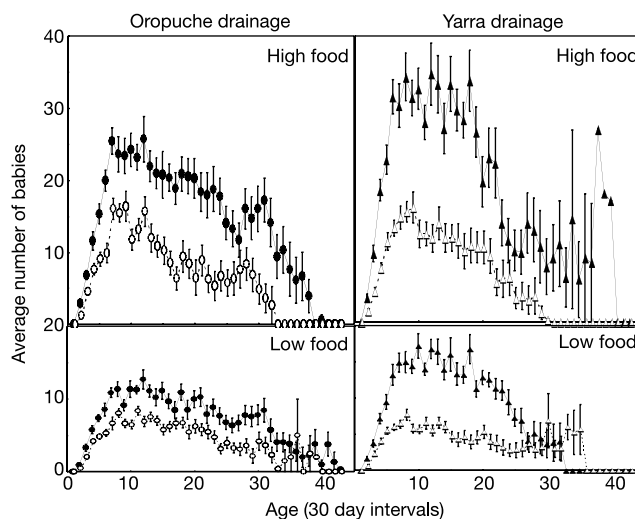


Figure 2 Fecundity. Number of offspring (*y*-axis) produced per 30 days (age, *x*-axis). We report the average number (± 1 s.e.m.) of babies produced by each treatment group per 30 days. The magnitude of the standard errors increases and fecundity becomes more erratic with increasing age because the sample sizes decline. Oropuche (circles), Yarra (triangles), high predation (filled), low predation (open).

summarize here results for only age at last reproduction (Fig. 1; for age at death see Supplementary Information).

First, we compared the fit of a Gompertz versus an exponential function to the data (S-Plus, CensorReg routine)²⁴. The Gompertz function provided a significantly better fit to the data in all comparisons (likelihood ratio test, $P < 10^{-4}$; Supplementary Information) indicating that mortality rate increases with age in all treatment groups.

Initial analyses of the age at last reproduction, assuming a Gompertz distribution, showed a significant interaction between drainage and predation, so we analysed drainages separately. High and low predation sites differ significantly in both drainages (Oropuche: z -value = -2.76 , $P = 0.0057$; Yarra: z -value = -3.23 , $P = 0.0012$). There was also a significant effect of food in the Oropuche drainage ($z = 2.00$, $P = 0.0459$) and a significant food by predation interaction in the Yarra drainage ($z = 2.61$, $P = 0.0091$). Food effects reflect the longer reproductive life spans of low food fish in the high and low predation sites from the Oropuche drainage and the low predation site from the Yarra drainage (Table 1).

The rate of ageing¹³ is lower in the high predation localities in all four paired comparisons (drainage by food) between guppies from high and low predation localities (Table 1). This index is a function of the α and γ parameters of the Gompertz equation [$m_x = \alpha e^{\gamma x}$] and hence of what is interpreted as the baseline mortality rate and the rate of increase in mortality with age²⁵. This result reflects the lower rates of cessation of reproduction throughout the lives of the high predation fish (Fig. 1, details and alternative indices of senescence in Supplementary Information).

Guppies from high predation localities begin reproduction at an earlier age²⁰, cease reproduction at a later age (Fig. 1) and hence have a longer reproductive lifespan. They also have longer total life spans (Table 1) but there are no differences among treatment groups in post-reproductive lifespan (data not shown). The differences in lifespan are thus concentrated in the duration of reproduction. These results do not conform with the conventional prediction that the populations from low mortality rate environments will also evolve delayed senescence if senescence is assayed as reproductive lifespan or rate of ageing¹³. We obtained similar results for guppies from a third pair of populations in a small pilot study that preceded this experiment (Supplementary Information).

A second way of evaluating senescence is through changes in the rate of production of offspring with age. Guppies from high predation localities in both drainages sustained a higher rate of

production of offspring throughout their lives (Fig. 2). Litter size is positively correlated with female size, so offspring production per unit time increased for seven to eight months after maturity in association with continued growth of the mother, then levelled off as females attained asymptotic body size. All treatment groups then experienced a progressive decline in offspring production with age, which is consistent with reproductive senescence.

The distribution of age-specific fecundity, which is the number of offspring produced per unit time, is triangular in shape and can be described with a modification of the fecundity function suggested by McMillan *et al.*²⁶. The key parameters are the age at which fecundity is greatest ($D_{\max}$) then the rate of decline in fecundity (b). The age at maximum fecundity was significantly later in low food treatments and in fish from the Oropuche drainage (see Methods). There was a non-significant trend ($P = 0.107$) for guppies from high predation communities to attain peak fecundity at a later age than those from low predation communities. The rate of decline in fecundity with age is only affected significantly by drainage ($P < 10^{-4}$); it is steeper in the Yarra drainage (Table 1). We obtained similar results with an alternative analysis in which we fit a linear regression to fecundity on age for ages beyond $D_{\max}$ (b_{lin} , Table 1) (Supplementary Information). Although not significant, the rate of decline in fecundity estimated as b_{lin} is always lower in guppies from high predation localities. Whereas these analyses do not provide statistical support for differences among predator communities in reproductive senescence, they also do not comply with the prediction for delayed senescence in guppies from low predation communities. If anything, the trends are in the opposite direction. A similar analysis of reproductive value, which is a composite measure of age-specific mortality and reproduction, yields qualitatively similar results (Supplementary Information).

Finally, we used the fast start escape response as an assay of neuromuscular performance. This response is a stereotyped burst of movement used to evade a striking predator. Peak accelerations during fast starts can be up to 40 body lengths s^{-2} (ref. 27). Age-related declines in the neuromuscular system have been documented as a cause of reduced physical performance²⁸. Because the fast start reflects a fish's neuromuscular performance, it serves as an index of physiological senescence. We estimated the maximum acceleration for a subset of our fish at an age of approximately 12 months then again at an age of approximately 26 months. There was a significant deceleration in all fish as they aged ($F_{1,35} = 36.85$, $P < 0.0001$), and a significant interaction between age and predation ($F_{1,35} = 5.6$, $P = 0.0237$) because guppies from high predation

Table 1 Life history statistics by treatment group and parameter estimates for the triangular fecundity function

Drainage	Food	Predation	Total life span (days)*	Age at last reproduction (days)*	Rate of ageing ¹³ (ω)	Number of litters*	Number of offspring*	Estimated parameters of the triangular fecundity function and b_{lin}					
								$D_{\max}$	M_{∞}	k	t_0	b	b_{lin}
Oropuche	High	High	1006.6 (636–1,264)	883.2 (533–1,216)	0.0047	27.7 (14–43)	545.4 (204–1,114)	10.42	3.61	0.47	1.07	0.010	–0.026
		Low	746.4 (338–1,136)	615.3 (251–976)	0.0117	16.8 (5–32)	217.6 (18–568)	9.57	4.17	0.46	0.99	0.015	–0.045
	Low	High	1030.9 (434–1,464)	914.9 (340–1,294)	0.0048	27.6 (7–38)	258.8 (51–444)	10.52	2.69	0.50	1.42	0.010	–0.025
		Low	844 (461–1,230)	730.4 (255–1,181)	0.0083	20.5 (5–29)	138.3 (17–212)	10.78	3.31	0.39	0.99	0.015	–0.037
Yarra	High	High	850.8 (448–1,228)	724.9 (324–1,228)	0.0100	25.6 (11–43)	607.7 (169–1,114)	9.12	2.12	0.30	0.79	0.039	–0.039
		Low	699.4 (164–1,291)	553.3 (115–908)	0.0131	15.7 (2–32)	204.6 (6–476)	8.59	3.73	0.41	1.07	0.035	–0.075
	Low	High	803 (151–1,075)	728.2 (109–976)	0.0035	25.1 (2–35)	283.5 (5–467)	10.78	2.49	0.27	0.67	0.027	–0.030
		Low	788.2 (396–1,253)	667.6 (306–1,066)	0.0110	18.5 (8–31)	115.0 (9–278)	9.09	1.29	0.33	0.79	0.029	–0.044

$D_{\max}$ equals the time interval in which fecundity was greatest. M_{∞} = the potential maximum daily fecundity (log transformed). t_0 = the first day of offspring production. k characterizes the rate of increase in fecundity. b and b_{lin} characterize the rate of decrease in fecundity with age.

* mean (range).

environments experience a more rapid deterioration in physiological performance with age than do their counterparts from low predation environments (Fig. 3). In separate analyses of each age, high predation guppies were significantly faster than low predation guppies when they were young ($F_{1,35} = 5.64$, $P = 0.0231$) but the two groups did not differ significantly when they were old ($F_{1,35} = 0.66$, $P = 0.42$).

Our results do not comply with the classical Medawar–Williams theory when senescence is evaluated in terms of survival, fecundity or reproductive value. Guppies from high predation localities have lower rates of ageing and do not differ in reproductive senescence relative to those from low predation localities. The only aspect of our results that complies with predictions is our assay of performance. Why did the unexpected happen? First, prior empirical support for these predictions is limited. In support of the classical theory, research on *Daphnia*¹⁴, opossums⁹ and grasshoppers¹² all found that populations with lower mortality rates had delayed senescence. In addition, Stearns *et al.*¹⁵ successfully selected for lower intrinsic mortality rates in *Drosophila melanogaster* that experienced lower extrinsic adult mortality rates.

There are three exceptions to the classic predictions based on extrinsic mortality that can contribute to our unexpected results. Williams proposed that “senescence should be more rapid in those organisms that do not increase markedly in fecundity after maturity than those that do show such an increase”². If fecundity increases with age, then this increase can offset the age-specific decline in fitness caused by mortality. In nature, guppies from high predation localities grow faster and attain larger asymptotic body sizes than those from low predation localities²³. Because litter size is directly proportional to body size, their faster growth translates into a higher rate of increase in fecundity with age (Supplementary Information). A consequence of this increase in fecundity is that it should offset some of the differences in mortality rate between high and low predation localities and hence result in smaller differences in senescence than expected from mortality rate alone²⁹.

Charlesworth⁴ and Abrams⁵ show that how organisms evolve in response to extrinsic mortality depends upon whether or not populations are subject to density regulation. To paraphrase one prediction, predators increase mortality rate by eating prey, but may decrease mortality rates indirectly by reducing density and increas-

ing per capita resource availability. If older age classes benefit more than younger age classes from higher resource availability, then higher mortality can cause the evolution of delayed senescence, even though increased mortality without an indirect effect of density predicts the evolution of earlier senescence. A diversity of other responses is possible, depending on the effects of density on the survival of different age classes and fecundity. Such indirect effects of predation occur in natural populations of guppies²³. Furthermore, the significant interaction between food availability and predation in the Yarra replicate argues that resource availability can affect the evolution of senescence; however, we do not yet have sufficient knowledge of the nature of density regulation in guppies to formally test these theories. There has already been a definitive demonstration in *Drosophila melanogaster* that ecology can alter the evolution of senescence. Luckinbill and Clare³⁰ showed that selection on late-life reproductive success causes the evolution of later senescence if larval density is high, but has no effect on the evolution of senescence if larval density is low. Stearns *et al.*¹⁵ successfully selected for later senescence and the evolution of other life history traits by decreasing adult mortality rates, but only after increasing larval density and decreasing food supply.

Abrams⁵, and Williams and Day⁶ present a third alternative by considering that the risk of mortality due to factors like disease or predation is often dependent on the condition of the individual. For example, if an age-related decline in escape response (physiological senescence) increases the risk of predation, then predation will select for improved ability to evade predators. A consequence of selection for improved escape performance will be deferred senescence in the age classes that have high reproductive potential. The strength of such selection should decline as an individual's reproductive potential declines with age. At the same time, the costs of deferred senescence will accumulate. Predation can thus select for a lower rate of senescence early in life but accelerated senescence late in life⁶. Our finding that high predation guppies have faster rapid-start responses early in life suggests that predators select for improved escape performance. The predicted lower rate of senescence early in life but more rapid senescence later in life is consistent with the trend in both drainages for high predation guppies to have a delayed, but then more rapid, acceleration in the cessation of reproduction and mortality (Fig. 1, Supplementary Information) relative to low predation guppies.

All three alternatives have properties that are applicable to guppies and all may contribute to our unexpected results. More generally, the classical prediction that high extrinsic mortality will cause the evolution of earlier senescence has dominated our thinking because it makes intuitive sense and has not been challenged by discordant empirical data. Our results provide new incentive to consider the importance of the derived models for the evolution of senescence. □

Methods

Localities and lab rearing: adult females were collected from the Yarra and Oropuche drainages in April, 1998. The approximate grid references for each collection, read from 1:25,000 topographic maps, are: Yarra River, PS940(N) PS804(W); Yarra Tributary (Limon River), PS876(N) PS834(W); Oropuche River, PS788(N) QS043 (W); Oropuche Tributary (Campo River), PS813 (N) PS971(W). The first generation of laboratory-born offspring were born between May and July, 1998, then were reared to adulthood and mated to produce the second generation of laboratory born offspring in November, 1998. Ten young produced by these crosses were reared in two groups of five in 8 l aquaria to an age of 25–30 days, then were sexed and two females were selected from each litter for inclusion in the senescence assay (initiated in January and February 1999). Females were reared by themselves in 8 l aquaria and were mated beginning one week after entering the assay—five weeks before the minimum age at first parturition. All other aspects of rearing were as in our previous studies. Survival at all stages was generally >95%, which is part of the basis for arguing that there is little or no inadvertent selection associated with this breeding design.

Statistics

To calculate survival we used the Censor Reg routine in S-Plus with the extreme value distribution²⁴. The evaluation of the Gompertz function and the associated slopes and

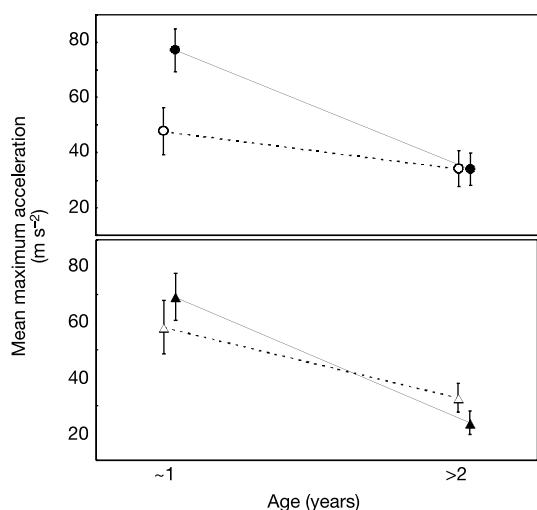


Figure 3 Maximum acceleration. The average maximum acceleration (± 1 s.e.m.) during the induced alarm response (y-axis) is plotted against age (x-axis) with the results for each drainage presented separately. Oropuche high predation (filled circles), Oropuche low predation (open circles), Yarra high predation (filled triangles), Yarra low predation (open circles).

intercepts of the fitted curves can be obtained as functions of the extreme value distribution as described in the S-Plus Guide to Statistics²⁴.

Fecundity was calculated using our modification of the triangular fecundity function²⁶, takes the form:

$$\ln(M_x) = M_\infty (1 - e^{-(kx+t_0)}) e^{-bx} \quad (1)$$

where M_x is the fecundity at age x , and the parameters M_∞ , k , t_0 and b were fitted by minimizing the sums of squares. M_∞ is the potential maximum daily fecundity (log transformed). t_0 is the first time period of offspring production, where each unit of time equals 30 days. k characterizes the rate of increase in fecundity at age x whereas b characterizes the rate of decrease in fecundity at age x . At early ages the term $(1 - e^{-(kx+t_0)})$ dominates and accounts for the rise in fecundity, whereas at later ages the term e^{-bx} dominates and largely determines the rate of decline in fecundity. The age at which fecundity is greatest, $D_{\max}$, is given by

$$D_{\max} = \frac{1}{k} \left(\ln \left(\frac{b+k}{b} \right) + t_0 \right) \quad (2)$$

Although b is a measure of the decline in fecundity, its value is also influenced by the increasing phase of the fecundity curve. A second measure of the rate of decline that is independent of the increasing phase is a linear regression of fecundity on age for ages beyond $D_{\max}$; we refer to the slope of this regression as b_{lin} (Supplementary Information). We used $D_{\max}$ for statistical comparisons among treatment groups of the age at maximum fecundity and b and b_{lin} for the rate of decline in fecundity with age. Both b and b_{lin} differed significantly among drainages; b_{lin} also differed among food levels. Both categories of variables can serve as indices of senescence. Standard errors for the estimates were obtained by the delete-one jackknife³¹ (Supplementary Information). To remove heteroscedasticity we used the rank-transformed values of b . The pseudovalues from the jackknife procedure were used to test for variation due to drainage, food ration, predation type and all interactions.

Performance was evaluated for 9–13 fish from each population at both ages. Food effects were not significant after performance was size corrected and was not included in the analysis. Fish were placed in a glass tank with a 1 cm² reference grid on the bottom. This focal tank was in turn placed within a larger glass tank that helped regulate water temperature. Fast-starts were elicited then filmed at 500 frames per sec (Redlake Motionscope camera) and recorded to VHS videocassette tape. Recorded sequences were converted from analogue (VHS) to digital format and saved as AVI files. The dorsal midline of the first three tail beat cycles were digitized using a modification of the public domain NIH Image program (available at <http://rsb.info.nih.gov/ni-image/>) for the Apple Macintosh. We fit a cubic spline function to each digitized midline within a sequence and used this function to find the coordinates of an unmeasured landmark at 0.35 total length from the tip of the head, which we used as a proxy for the centre of mass. We used a MSE quintic spline algorithm to smooth the displacement versus time data and to estimate instantaneous velocities and accelerations throughout a digitized sequence. From the fitted functions, we computed maximum accelerations occurring within the first 22 ms. The analysis was a repeated measures Anova with predators, drainage and time as main effects. Drainage and interactions between drainage and other variables were not significant. Maximum acceleration was the dependent variable. See ref. 32 for additional details.

Received 16 April; accepted 16 August 2004; doi:10.1038/nature02936.

- Medawar, P. B. *An Unsolved Problem of Biology* (H. K. Lewis and Co., London, 1952).
- Williams, G. C. Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
- Reznick, D. A., Bryga, H. & Endler, J. A. Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
- Charlesworth, B. *Evolution in Age Structured Populations* (Cambridge Univ. Press, Cambridge, 1980).
- Abrams, P. Does increased mortality favor the evolution of more rapid senescence? *Evolution* **47**, 877–887 (1993).
- Williams, P. D. & Day, T. Antagonistic pleiotropy, mortality source interactions and the evolutionary theory of senescence. *Evolution* **57**, 1478–1488 (2003).
- Austad, S. N. & Fischer, K. E. Mammalian aging, metabolism, and ecology—evidence from the Bats and Marsupials. *J. Gerontol.* **46**, B47–B53 (1991).
- Promislow, D. E. L. Senescence in natural-populations of Mammals—a comparative study. *Evolution* **45**, 1869–1887 (1991).
- Austad, S. N. Retarded senescence in an insular population of Virginia opossums (*Didelphis virginiana*). *J. Zool.* **299**, 695–708 (1993).
- Holmes, D. J. & Austad, S. N. The evolution of avian senescence patterns—implications for understanding primary aging processes. *Am. Zool.* **35**, 307–317 (1995).
- Keller, L. & Genoud, M. Extraordinary lifespans in ants: a test of evolutionary theories of ageing. *Nature* **389**, 958–960 (1997).
- Tatar, M., Gray, D. W. & Carey, J. R. Altitudinal variation for senescence in *Melanoplus* grasshoppers. *Oecologia* **111**, 357–364 (1997).
- Ricklefs, R. E. Evolutionary theories of aging: confirmation of a fundamental prediction, with implications for the genetic basis and evolution of life span. *Am. Nat.* **152**, 24–44 (1998).
- Dudykha, J. & Tessier, A. Natural genetic variation of life span, reproduction, and juvenile growth in *Daphnia*. *Evolution* **53**, 1744–1756 (1999).
- Stearns, S. C., Ackermann, M., Doebl, M. & Kaiser, M. Experimental evolution of aging, growth, and reproduction in fruitflies. *Proc. Natl Acad. Sci. USA* **97**, 3309–3313 (2000).
- Hendry, A. P., Morbey, Y. E., Berg, O. K. & Wenburg, J. K. Adaptive variation in senescence: reproductive lifespan in a wild salmon population. *Proc. R. Soc. Lond. B* **271**, 259–266 (2004).
- Kirkwood, T. B. L. & Austad, S. N. Why do we age? *Nature* **408**, 233–238 (2000).
- Partridge, L. & Gems, D. Mechanisms of ageing: public or private? *Nature Rev. Genet.* **3**, 165–175 (2002).

- Reznick, D. N., Butler, M. J. I., Rodd, F. H. & Ross, P. Life history evolution in guppies (*Poecilia reticulata*). 6. Differential mortality as a mechanism for natural selection. *Evolution* **50**, 1651–1660 (1996).
- Reznick, D. N., Buckwalter, G., Groff, J. & Elder, D. The evolution of senescence in natural populations of guppies (*Poecilia reticulata*): a comparative approach. *Exp. Gerontol.* **36**, 791–812 (2001).
- Carvalho, G. R., Shaw, P. W., Magurran, A. E. & Seghers, B. H. Marked genetic divergence revealed by allozymes among populations of the guppy *Poecilia reticulata* (Poeciliidae), in Trinidad. *Biol. J. Linn. Soc.* **42**, 389–405 (1991).
- Reznick, D. N. The impact of predation on life history evolution in Trinidadian guppies: the genetic components of observed life history differences. *Evolution* **36**, 1236–1250 (1982).
- Reznick, D. N., Butler, M. J. I. & Rodd, F. H. Life history evolution in guppies 7: The comparative ecology of high and low predation environments. *Am. Nat.* **157**, 126–140 (2001).
- S-Plus. *S-Plus 6 for Windows Guide to Statistics* Vol. 3 (Insightful Corporation, Seattle, Washington, 2001).
- Finch, C. E., Pike, M. C. & Whitten, M. Slow mortality rate accelerations during aging in some animals approximate that of humans. *Science* **249**, 902–905 (1990).
- McMillan, I., Fitz-Earle, M. & Robson, D. S. Quantitative genetics of fertility. I. Lifetime egg production of *Drosophila melanogaster*. *Theor. Genet.* **65**, 349–353 (1970).
- Domenici, P. & Blake, R. W. The kinematics and performance of fish fast-start swimming. *J. Exp. Biol.* **200**, 1165–1178 (1997).
- Delbono, O. Neural control of aging skeletal muscle. *Aging Cell* **2**, 21–29 (2003).
- Reznick, D., Ghalambor, C. & Nunney, L. The evolution of senescence in fish. *Mech. Ageing Dev.* **123**, 773–789 (2002).
- Luckinbill, L. S. & Clare, M. J. Selection for life span in *Drosophila melanogaster*. *Heredity* **55**, 9–18 (1985).
- Oppenheimer, L., Capizzi, T. P. & Miwa, G. T. Application of the jackknife procedures to inter-experiment comparisons of parameter estimates for the Michaelis–Menten equation. *Biochem. J.* **197**, 721–729 (1981).
- Ghalambor, C., Walker, J. A. & Reznick, D. N. Constraints on adaptive evolution: The functional trade-off between reproduction and burst swimming performance in the guppy (*Poecilia reticulata*). *Am. Nat.* **164**, 38–50 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This research was supported by NSF Grant and by the Academic Senate of the University of California. We thank P. Abrams, A. Bronikowski, M. Clark and P. Williams for comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R. (david.reznick@ucr.edu).

Population density drives the local evolution of a threshold dimorphism

Joseph L. Tomkins & Gordon S. Brown

Division of Environmental and Evolutionary Biology, Sir Harold Mitchell Building, University of St Andrews, St Andrews, Fife KY16 9TH, UK

Evolution can favour more than one reproductive tactic among conspecifics of the same sex^{1,2}. Under the conditional evolutionarily stable strategy, individuals adopt the tactic that generates the highest fitness return for their status: large males guard females, whereas small males sneak copulations^{3,4}. Tactics change at the status at which fitness benefits switch from favouring one tactic to favouring the alternative^{1,5}. This ‘switch-point’ is expressed in many species as a threshold between divergent morphologies³. Environmental and demographic parameters that influence the relative fitness of male tactics are predicted to determine a population’s switchpoint^{1,5} and consequently whether the population is monomorphic or dimorphic. Here we show threshold evolution in the forceps dimorphism of the European earwig *Forficula auricularia* and document the transition from completely monomorphic to classical male-dimorphic populations over a distance of only 40 km. Because the superior fighting ability of the dominant morph⁶ will be more frequently rewarded at high encounter rates, population density is likely to be a key determinant of the relative fitness of the

alternative tactics, and consequently the threshold. We show that, as predicted, population density correlates strongly with the shift in threshold, and that this factor drives the local evolution of the male dimorphism in these island populations. Our data provide evidence for the origin of phenotypic diversity within populations^{7–9}, through the evolution of a switchpoint in a conditional strategy that has responded to local population density.

The morphological adaptations of males using alternative tactics under the conditional evolutionarily stable strategy (ESS) represent an extreme case of phenotypic plasticity⁹. In the European earwig *F. auricularia*, there has been a longstanding interest in this plasticity, in which the alternative male tactics are characterized by a dimorphism in male forceps length^{10–12} (Fig. 1a, b). The dimorphism is characterized by a large-bodied ‘macrolabial’ morph with long forceps, and a small ‘brachylabial’ morph with short forceps. Male *F. auricularia* of both morphs use their forceps in courtship¹³ and fighting^{6,14} in which macrolabial males are

competitively superior⁶. Data that quantify sexual selection in the field show that macrolabial males are more likely to be found guarding a female than are brachylabial males (J.L.T., unpublished data). This is consistent with behavioural roles in which macrolabial males guard females and brachylabial males are forced to sneak copulations. The male dimorphism led early researchers to consider that the macrolabial males were a different species, *Forficula forcipata*¹¹; however, ‘common-garden’ rearing experiments show that, within populations, the dimorphism is largely dependent on the nutrition of the developing nymph¹². Subtle between-population variation in morphological threshold has been documented in a number of species reared under common-garden conditions, including Farne island populations of *F. auricularia*¹², the mite *Sancassania berleset*¹⁵ and the dung beetle *Onthophagus taurus*¹⁶. These studies suggest that differences between populations represent the genetic divergence of the ESS switchpoints between populations. Artificial selection experiments support the notion that thresholds can harbour large amounts of genetic variation^{17–19}, fuelling population divergence.

British island populations of *F. auricularia* have a significantly higher proportion of the macrolabial male morph than do mainland populations (Fig. 2a). This phenomenon is equally true in the northeast of the British Isles as in the southwest, suggesting that the variation between islands and the mainland is not simply due to gross variation in climate or to milder microclimatic conditions on islands (air temperature at Boulmer in Northumberland is on average 1.5 °C cooler than in St Mawgan in Cornwall; Met office data). To understand why island populations of earwigs are male-dimorphic we concentrated on population variation among North Sea islands in the Firth of Forth and the Farne Islands. This minimized differences in the climate, geological history, habitat and the ecology and diversity of interacting species. Only these islands are considered in further analyses. Earwig populations in the Firth of Forth varied considerably in the observed ratio of macrolabial to brachylabial males, ranging from 0–20% and in the Farne group from 8–45% (Fig. 2).

Variation in morph frequency is due to a change in position of the dimorphic threshold relative to the population mean body size. Hence, morph ratio variation can either be due to threshold evolution independent of mean body size, or to changes in mean body size around a static threshold, or both. The former (shown in Fig. 1a) is consistent with ESS theory under which the threshold is determined by population-specific variation in fitness functions¹; the latter is likely when a single population is reared across an environmental gradient¹². To determine whether populations varied significantly in the position of the body size threshold, we categorized males as macrolabial or brachylabial (see Methods), then performed a logistic regression with morph as the dependent ‘variable, pronotum width (a measure of body size) as a covariate and island as a factor (including all island populations with at least one macrolabial male). The whole model was significant ($\chi^2_{21} = 1,456$, $P < 0.001$), as was the effect of pronotum width, indicating, as expected, that larger individuals are more likely to be macrolabial ($\chi^2_1 = 1,134$, $P < 0.001$); the population term was also significant ($\chi^2_{20} = 154$, $P < 0.001$), demonstrating that populations differ in the absolute position of the morphological threshold (Fig. 3a, b).

Further evidence that the position of the threshold is not fixed comes from the positive correlation between a population’s mean pronotum width and the absolute position of the morphological threshold in the population ($r_s = 0.77$, $n = 21$, $P = 0.001$); the ESS switchpoint apparently having evolved to some extent in parallel with increasing body size.

A logistic regression model, with pronotum width standardized within each population to have a mean of zero and a standard deviation of one, revealed that populations also differed in the position of the threshold relative to the population mean (population: $\chi^2_{20} = 508$, $P < 0.001$). Hence, not only does the absolute

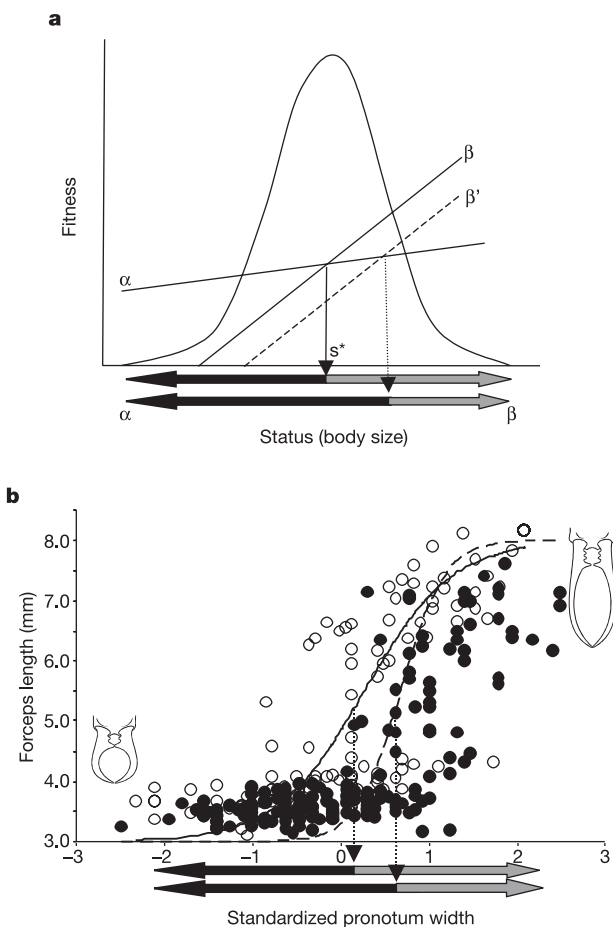


Figure 1 The status-dependent ESS and threshold variation in *F. auricularia*. **a**, The conditional ESS model with status-dependent alternative tactics¹. Fitness functions of the tactics α and β are status-dependent. Individuals adopt the tactic from which they derive the highest fitness; hence the proportion of males of each morph in the population is determined by the position of the intersection of the fitness functions: the ESS switchpoint s^* . Environmental and demographic variables that affect the fitness functions (for example, a decline in density shifting β to β') generate population variation in s^* and hence between-population variation in morph ratio¹ (shaded arrows). **b**, The nonlinear relationship between male body size and forceps length that characterizes male dimorphic populations of *F. auricularia*. Inset are brachylabial and macrolabial forceps. Two populations are shown: Bass Rock (open circles) and Knoxes Reef (closed circles). Cubic splines²⁷ summarize the male dimorphism in each population. The ESS s^* is reflected by the morphological threshold between the morphs. Shifts in ESS s^* due to changes in the fitness functions of the morphs (for example, **a**) are responsible for shifts in the morphological threshold between populations.

body size threshold show variation between populations, but so too does the relative position of the threshold: the parameter that determines male morph ratio. Population variation in the morphological threshold and body size has occurred on an extremely small scale. For example, the islands of East Wideopen (EWO) and Knoxes Reef (KR) in the inner group of Farne islands are separated by just 400 m of sea, and are both connected by causeways to the same island at low tide. Nonetheless, populations of earwigs on these adjacent islands differ significantly in both size (pronotum width mean $\pm$ s.e.m: EWO = 2.05 ± 0.01 , KR = 1.88 ± 0.01 ; $t_{384} = 13.5$, $P < 0.001$) and the absolute and relative position of the threshold (final model, $\chi^2_2 = 209$, $P < 0.001$; pronotum width, $\chi^2_1 = 196$, $P < 0.001$; island, $\chi^2_1 = 26.7$, $P < 0.001$; using standardized pronotum width, island = $\chi^2_1 = 19.1$, $P < 0.001$ Fig. 3c).

Theoretically, the increase in frequency of macrolabial males on island populations occurs because the fitness function of macrolabial males is elevated relative to that of brachylabial males^{1,5}. Male earwigs guard females in chambers under stones and driftwood; following oviposition, the female will not re-mate and expels the male. In *F. auricularia*, macrolabial males have an advantage over brachylabial males when competing for females (unpublished data)⁶. Nevertheless, adaptations for mate guarding, such as elongate forceps, are likely to be advantageous only where challenges occur regularly. The intensity of sexual selection is predicted to vary with certain demographic and ecological parameters, such as density, the spatial distribution of resources and the operational sex ratio (OSR)^{20,21}. In *F. auricularia*, the probability of being challenged by an unpaired male, or conversely of successfully usurping a guarding male, depends on the probability of encountering another burrow. Across the very similar island habitats studied here, the predominant factor determining encounter rate is likely to be population density. Where earwigs occur at high densities, challenges over guarded females are likely to be frequent, elevating the fitness of males with adaptations for fighting for or defending females. Extremely high population densities of earwigs are a conspicuous

feature of the ecology of the islands of the Firth of Forth and Farnes. Across the North Sea island populations, we found that there was a significant relationship between the overall density of the earwigs and the proportion of macrolabial males in the population ($r_{s22} = 0.66$, $P = 0.001$, Fig. 4).

OSR is an important factor determining the intensity of sexual selection²⁰. Although the sex ratio did become more male biased as population densities increased across islands, there was no significant relationship with male morph ratio (see Supplementary Information).

Threshold evolution in earwigs on these North Sea islands has taken place in similar climatic conditions and similar habitats. The extraordinarily high population densities that seem to be driving threshold evolution may arise due to the escape from parasitic species or competitors present on the mainland. This hypothesis predicts increasingly high densities of earwigs as islands become more difficult for other species to colonize. There was a significant increase in earwig density with increasing distance of the island from the mainland ($r_{s22} = 0.616$, $P = 0.002$), but no increase in density with decreasing island size ($r_{s22} = -0.05$, $P = 0.836$). Despite the increase in density with distance to the mainland, there was no increase in the proportion of macrolabial males with distance to the mainland ($r_{s22} = 0.36$, $P = 0.101$).

The input of organic material on which earwigs feed may also be an important determinant of population density on small islands with sparse vegetation. Earwigs eat both guano and the carcasses of dead seabirds. Nevertheless, the biomass of ground-nesting birds per square metre on each island did not correlate with earwig density ($r_{s21} = 0.02$, $P = 0.915$).

We have shown that threshold evolution can occur on an extremely small geographic scale in a pattern consistent with local variation in demographic parameters that are theoretically important to variation in the intensity of sexual selection²¹. Male-dimorphic threshold traits have been shown to respond readily to selection^{17,19}, and although population variation consistent with

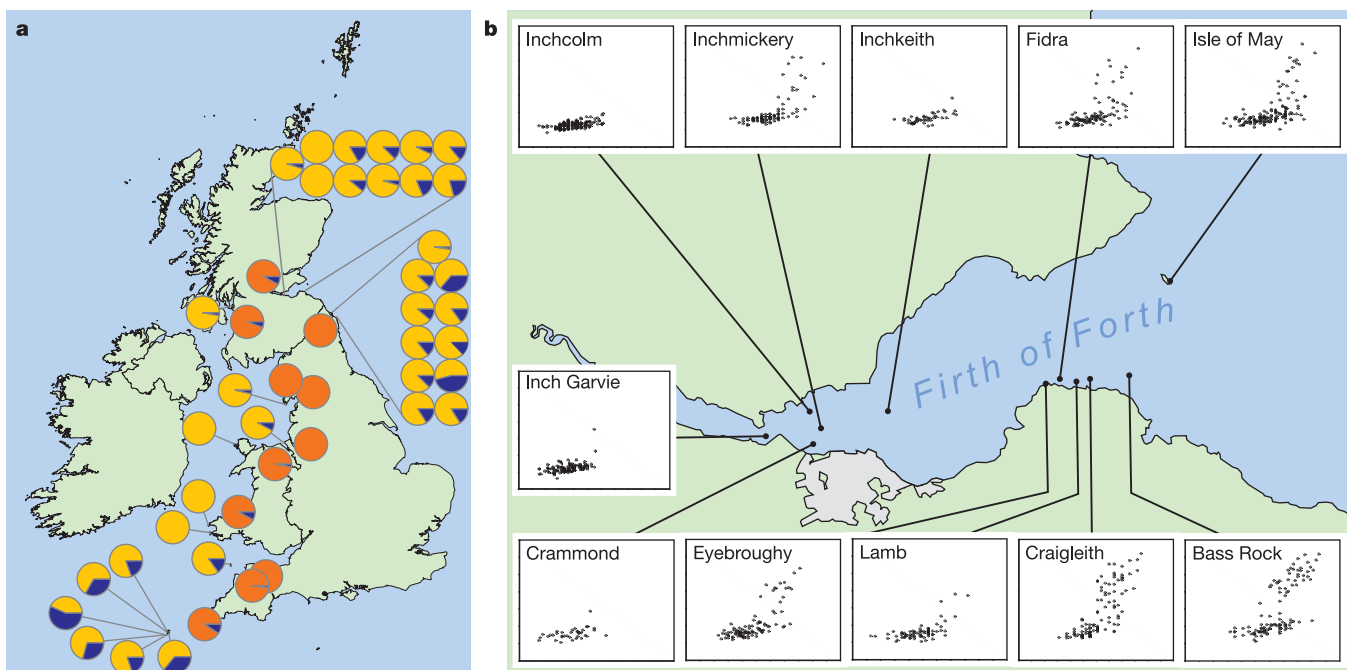


Figure 2 Morph ratio variation among populations of *F. auricularia*. **a**, Map of Britain and Ireland with pie charts showing the proportion of male morphs in island and mainland populations of the European earwig *F. auricularia*. Macrolabial males represent a greater proportion of males in island populations than mainland populations (t -test on arcsine square-root transformed proportion macrolabial, $t_{44} = 3.17$, $P = 0.003$). Macrolabial

males (blue); brachylabial males in mainland populations (orange) and in-island populations (yellow). **b**, Variation between island populations in the Firth of Forth. Inset scatter plots show the relationship between forceps length (y axis, all graphs 2.0–9.0 mm) and pronotum width (x axis, all graphs 1.4–2.5 mm).

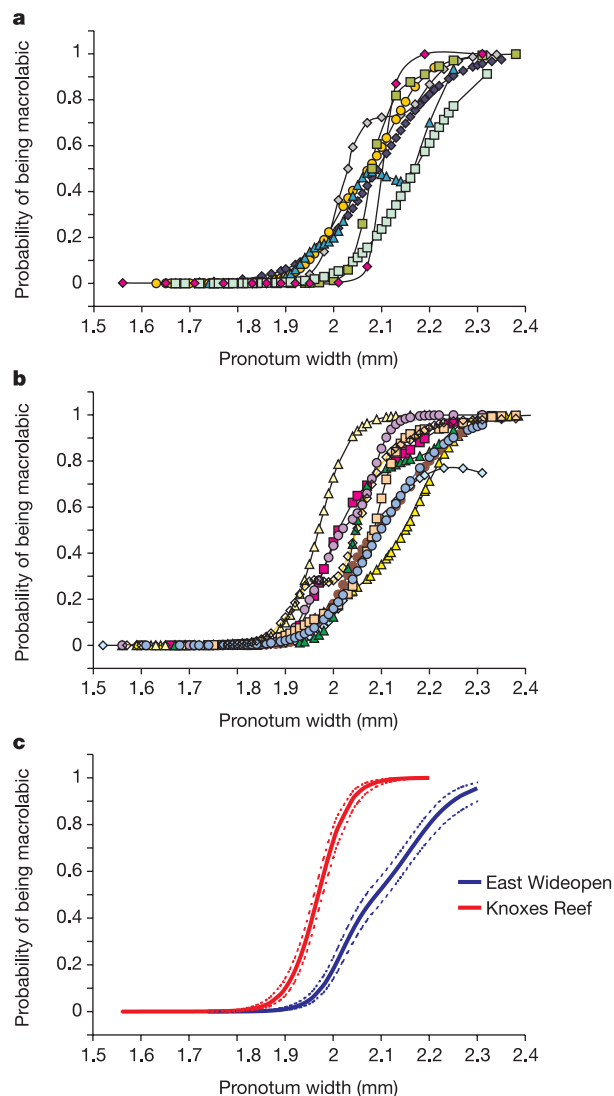


Figure 3 Variation in threshold among island populations of *F. auricularia*. **a**, **b**, Cubic splines summarizing the threshold variation in the reaction norm of forceps length on pronotum width for populations of *F. auricularia* in the Firth of Forth (**a**) and the Farne Islands (**b**). **c**, Population variation in threshold on a very small geographic scale: Knoxes Reef and East Wideopen are islands in the Farnes group and are about 400 m apart. Dashed lines are 95% confidence intervals for the splines.

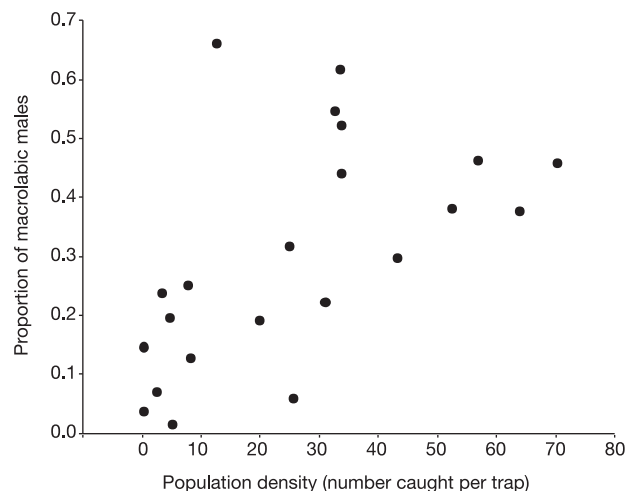


Figure 4 The relationship between density, estimated as the average number of earwigs caught in standard traps, and the proportion of macrolabial males in a population, for populations of *F. auricularia* from islands in the Firth of Forth and Farnes.

change in density has been proposed for dung beetles, the results were inconclusive²². Unlike previous studies, we have demonstrated that threshold evolution can tailor populations to fine-scale variation in demography. This is expected in threshold traits that reflect status-dependent alternative reproductive tactics, because any differences in the slopes or elevations of the fitness functions of alternative tactics affect the position of the ESS switchpoint (Fig. 1), thereby making it particularly sensitive to changes in parameters that alter relative fitness (see ref. 21 for an alternative perspective on the status-dependent ESS model). We are currently examining the extent to which gene flow between islands and with the mainland influences threshold variation among these populations.

We have documented a micro-evolutionary transition within a single species, over less than 40 km, from populations in which the ESS is for a single tactic expressed by all individuals, through to a classic conditional strategy in which alternative tactics are played in a status-dependent manner. These populations provide us with snap-shots in the evolution of a male dimorphism and support the game theoretic premise that ESSs are population specific^{1,2}. □

Methods

Collections

Earwig samples from the west coast and west-coast islands of Britain were collected from beneath driftwood and stones, or from the hollow stems of the hogweed *Heracleum sphondylium* (Umbelliferae). Collections were made between August and October 2002 (average sample size, 72 ± 9 ; range, 13–148). Trap sampling of the Farne Islands was carried out in August 2001 and of the Firth of Forth in September 2002 (average sample size, 121 ± 11 ; range, 37–227). Samples were stored at -20°C before measurement.

Measurements

Male earwigs from 22 focal islands in the Firth of Forth (Inchcolm, $n = 145$; Inch Garvie, $n = 100$; Crammond, $n = 37$; Inchmickery, $n = 92$; Inchkeith, $n = 145$; Eyebroughy, $n = 100$; Lamb, $n = 71$; Fidra, $n = 100$; Craigleith, $n = 100$; Bass Rock, $n = 105$; Isle of May, $n = 138$) and the Farne Islands (Lindisfarne, $n = 73$; Inner Farne, $n = 82$; Knoxes Reef, $n = 192$; West Wideopen, $n = 227$; East Wideopen, $n = 194$; Staple, $n = 224$; Brownsman, $n = 176$; South Wamses, $n = 135$; North Wamses, $n = 123$; Big Harcar, $n = 57$ and Longstone End, $n = 58$) were measured under a Leica MZ5 binocular microscope with an eyepiece graticule, or using Scion Image (NIH) software. Male pronotum width and right forceps length were measured. Male earwigs from west coast mainland and island populations were used only in the analysis of geographic variation in morph ratio. This analysis does not require precise knowledge of the morphological switch point and morphs were assigned by eye.

Switchpoint/threshold estimation

We have used the term 'threshold' to describe the morphological transition between morphs²³; the threshold estimates the position of the ESS 'switchpoint'¹. Thresholds were estimated using the method described in ref. 24 (hereafter referred to as the E&G method). We used a program written for this procedure in Splus 2000, in which 100 possible thresholds could be tested per population (K. Wilson). We chose the threshold at which the r^2 and the significance of the β_3 inflexion were maximized. Using the E&G method, the proportion of the population that lie each side of the morphological threshold was calculated as the proportion of a normal curve²⁵. This estimate of morph ratio was used to compare populations independent of among population variation in body size. Thus the E&G method was used to estimate the morphological threshold and population morph ratio. In addition, Kotiaho and Tomkins' (K&T) modification of the E&G method was used to assign earwigs to different morphs on the basis of their forceps length²⁶. These data were used in the logistic regression analyses of population variation in threshold. The K&T method was also used to summarize the relationship between forceps length and pronotum width using cubic splines (Fig. 3a–c). Output from the cubic spline program *glmSWIN*²⁷ was used to interpolate the body size at which males were macrolabial with a 50% probability; this K&T threshold correlated strongly with the E&G threshold ($r_{s22} = 0.88$, $P < 0.001$).

Density

Earwig density was estimated from trap catches. Earwigs aggregate in crevices during the day and traps were designed to exploit this behaviour. Traps were 150-ml cylindrical plastic vials containing a roll of standard corrugated cardboard (see Supplementary Information for further details). The average number of traps per island was 22 ± 2 . After 3 weeks we returned to the islands and collected the earwigs residing in the traps. Analysis of variance (ANOVA) with population as a random effect revealed that there was significant variance between islands in trap catches ($F_{21,502} = 9.57$, $P = 0.001$). Density data per island were left-skewed and on the verge of non-normality (Shapiro–Wilk = 0.91, d.f. = 21, $P = 0.05$) and because log transformation did not normalize the data nonparametric statistics were used.

Biomass of nesting birds

The number of pairs of the predominant species of ground-nesting birds nesting on each island of interest was calculated for the Farne Islands using the records of the National

Trust (1924–2003), and for the Firth of Forth using data from the Seabird Monitoring Program (1969–2003). The biomass of nesting birds per island was calculated using body weight data²⁸ (see Supplementary Information for exact details and species).

Island area and habitat

Island area was calculated from 1:25,000 maps obtained from the Ordnance Survey website and saved as JPEG files. Area within the mean high-water contour was measured using Scion Image. The islands are small; 17 out of the 22 have an area above the mean high-tide level, equivalent to a circle with a diameter of less than 270 m. This area will be an overestimate of the land habitable to earwigs because of spray, spring tides and storms. The islands are rock with a cap of shallow soil or peat, which is restricted to rock fissures on Longstone End, Eyebroughy and Big Harcar. Knoxes Reef is a shingle bank that has built up behind a reef of rock. The vegetation on all of the islands is dominated by grasses and annual herbs. Only Crammond and Lindisfarne support trees.

Received 20 May; accepted 3 August 2004; doi:10.1038/nature02918.

- Gross, M. R. Alternative reproductive tactics: diversity within sexes. *Trends Ecol. Evol.* **11**, 92–98 (1996).
- Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
- Hunt, J. & Simmons, L. W. Status-dependent selection in the dimorphic beetle *Onthophagus taurus*. *Proc. R. Soc. Lond. B* **268**, 2409–2414 (2001).
- Emlen, D. J. Alternative reproductive tactics and male dimorphism in the horned beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Behav. Ecol. Sociobiol.* **41**, 335–341 (1997).
- Hazel, W. N., Smock, R. & Johnson, M. D. A polygenic model for the evolution and maintenance of conditional strategies. *Proc. R. Soc. Lond. B* **242**, 181–187 (1990).
- Radesäter, T. & Halldórsdóttir, H. Two male types of the common earwig: male–male competition and mating success. *Ethology* **95**, 89–96 (1993).
- West-Eberhard, M. J. Alternative adaptations, speciation and phylogeny (a review). *Proc. Natl Acad. Sci. USA* **83**, 1388–1392 (1986).
- West-Eberhard, M. J. Phenotypic plasticity and the origins of diversity. *Annu. Rev. Ecol. Syst.* **20**, 249–278 (1989).
- West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford, New York, 2003).
- Huxley, J. S. Studies in heterotic growth (III). Discontinuous variation and heterogeneity in *Forficula*. *J. Genet.* **17**, 309–327 (1927).
- Bateson, W. & Brindley, H. H. On some cases of variation in secondary sexual characters, statistically examined. *Proc. R. Soc. Lond.* **1892**, 585–594 (1892).
- Tomkins, J. L. Environmental and genetic determinants of the male forceps length dimorphism in the European earwig *Forficula auricularia* L. *Behav. Ecol. Sociobiol.* **47**, 1–8 (1999).
- Tomkins, J. L. & Simmons, L. W. Female choice and manipulations of forceps size and symmetry in the European earwig *Forficula auricularia* L. *Anim. Behav.* **56**, 347–356 (1998).
- Forslund, P. Male competition and large size mating advantage in the European earwigs *Forficula auricularia*. *Anim. Behav.* **59**, 753–762 (2000).
- Tomkins, J. L., LeBas, N. R., Unrug, J. & Radwan, J. Testing the status-dependent ESS: population variation in fighter expression in the mite *Sancassania berlesii*. *J. Evol. Biol.* (in the press).
- Moczek, A. P., Hunt, J., Emlen, D. J. & Simmons, L. W. Threshold evolution in exotic populations of a polyphenic beetle. *Evol. Ecol. Res.* **4**, 587–601 (2002).
- Unrug, J., Tomkins, J. L. & Radwan, J. Alternative phenotypes and sexual selection: can dichotomous handicaps honestly signal quality? *Proc. R. Soc. Lond. B* **271**, 1401–1406 (2004).
- Roff, D. A. *Evolutionary Quantitative Genetics* (Chapman and Hall, New York, 1997).
- Emlen, D. J. Artificial selection on horn body-length size allometry in the horned beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Evolution* **50**, 1219–1230 (1996).
- Emlen, S. T. & Oring, L. W. Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215–223 (1977).
- Shuster, S. M. & Wade, M. J. *Mating Systems and Strategies* (Princeton, Princeton, 2003).
- Moczek, A. P. The behavioral ecology of threshold evolution in a polyphenic beetle. *Behav. Ecol.* **14**, 841–854 (2004).
- Roff, D. A. The evolution of threshold traits in animals. *Q. Rev. Biol.* **71**, 3–35 (1996).
- Eberhard, W. G. & Gutierrez, E. E. Male dimorphisms in beetles and earwigs and the question of developmental constraints. *Evolution* **45**, 18–28 (1991).
- Zar, J. H. *Biostatistical Analysis* (Prentice-Hall, New Jersey, 1984).
- Kotiaho, J. S. & Tomkins, J. L. The discrimination of alternative male morphologies. *Behav. Ecol.* **12**, 553–557 (2001).
- Schluter, D. Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 846–861 (1988).
- Snow, D. W. & Perrins, C. M. *Birds of the Western Palearctic* Concise edition (OUP, Oxford, 1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Walton, J. Stewart-Clark, H. Hamilton Dalrymple, C. Agasim-Pereira, P. Barry, K. Smith, C. Gallagher, R. Selley, B. Sampson, J. Love, D. Mawer, M. Gurr, R. Dorrien-Smith, I. Bullock, S. Avery, J. Brown and G. Thompson for permission to visit the islands; R. Mavor and D. Jones for data on seabirds; island wardens, R. Summers, J. Wilson, I. Parkinson, A. Shrieve, R. Harvey, O. Gabb, M. Steele, B. Teunis and J. Thompson for logistical support; the boat crews, B. McConnell, A. Hall (SMRU1), S. Moss, J. Dale and N. Quick (SMRU2), B. Sheil and crew (Glad Tidings) and F. Mar (Sula); research assistants, A. Arthur, C. Benskin and J. Wernham; A. Arthur for artwork; K. Wilson for Splus 2000 code; N. LeBas for field assistance, discussion of the manuscript and insights into the data; and J. Alcock, N. Colegrave, D. Hosken, R. Knell, J. Kotiaho, S. Tomkins, T. Tregenza, J. Radwan, M. Ritchie, L. Simmons and K. Wilson for comments. J.L.T. is supported by a BBSRC David Phillips Research Fellowship. G.S.B. is supported by a NERC PhD studentship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L.T. (jlt1@st-andrews.ac.uk).

Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex

Sabrina D. Dyall^{1,2}, Weihong Yan^{1,3}, Maria G. Delgadillo-Correa¹, Adam Lunceford³, Joseph A. Loo⁴, Catherine F. Clarke³ & Patricia J. Johnson^{1,2}

¹Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, 1602 Molecular Sciences Building, 609 Charles E. Young Drive East, Los Angeles, California 90095-1489, USA

²NASA Astrobiology Institute, Center for Astrobiology/IGPP, 3845 Slichter Hall, Los Angeles, California 90095-1567, USA

³Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, California 90095-1569, USA

⁴Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, 402 Paul D. Boyer Hall, Los Angeles, California 90095-1737, USA

Trichomonas vaginalis is a unicellular microaerophilic eukaryote that lacks mitochondria yet contains an alternative organelle, the hydrogenosome, involved in pyruvate metabolism. Pathways between the two organelles differ substantially: in hydrogenosomes, pyruvate oxidation is catalysed by pyruvate:ferredoxin oxidoreductase (PFOR), with electrons donated to an [Fe]-hydrogenase which produces hydrogen. ATP is generated exclusively by substrate-level phosphorylation in hydrogenosomes, as opposed to oxidative phosphorylation in mitochondria¹. PFOR and hydrogenase are found in eubacteria and amitochondriate eukaryotes, but not in typical mitochondria^{2–4}. Analyses of mitochondrial genomes indicate that mitochondria have a single endosymbiotic origin from an α -proteobacterial-type progenitor⁵. The absence of a genome in trichomonad hydrogenosomes⁶ precludes such comparisons, leaving the endosymbiotic history of this organelle unclear⁷. Although phylogenetic reconstructions of a few proteins indicate that trichomonad hydrogenosomes share a common origin with mitochondria^{8–11}, others do not^{2–4,7}. Here we describe a novel NADH dehydrogenase module of respiratory complex I that is coupled to the central hydrogenosomal fermentative pathway to form a hydrogenosomal oxidoreductase complex that seems to function independently of quinones. Phylogenetic analyses of hydrogenosomal complex I-like proteins Ndh51 and Ndh24 reveal that neither has a common origin with mitochondrial homologues. These studies argue against a vertical origin of trichomonad hydrogenosomes from the proto-mitochondrial endosymbiont.

To establish the ancestry of *T. vaginalis* hydrogenosomes and investigate whether hydrogenosomes derive from the proto-mitochondrial endosymbiont, we have taken a proteomics approach. Two-dimensional polyacrylamide gel electrophoretic analyses of soluble hydrogenosomal fractions have revealed a 48-kDa protein, Ndh51 (NADH dehydrogenase, 51 kDa), which is 57% identical to the 51-kDa subunit of mitochondrial respiratory complex I and to the NuoF (NADH:ubiquinone oxidoreductase, chain F) subunit of proteobacterial complex I. Complex I catalyses electron transfer from NADH to quinone and generally translocates protons across bacterial or mitochondrial membranes, creating an electromotive force that drives ATP production¹². Proteobacterial complex I contains 14 subunits, NuoA to NuoN, organized in three functional domains: an NADH dehydrogenase, a quinone-binding module, and a proton translocation module¹² (Fig. 1a). In addition to these core subunits, mitochondrial complex I also comprises up to 32 subunits of non-endosymbiotic origin^{12,13}. NuoF, NuoE and NuoG form the NADH dehydrogenase module bearing

NADH-binding and FMN-binding sites and four Fe-S clusters^{12,13}. *T. vaginalis* Ndh51 (Fig. 1b) has conserved putative NADH-binding and FMN-binding sites and four cysteine residues implicated in [4Fe-4S] cluster ligation¹³. We searched *T. vaginalis* databases with complex I subunit sequences from *Rickettsia prowazekii*, *Bos taurus* and *Arabidopsis thaliana*, including mitochondrion-specific sequences from the latter two species^{12,13}, and found potential homologues for NuoG and NuoE. Putative [Fe]-only hydrogenases³ were found as partial homologues to NuoG, with amino-terminal Fe-S cluster-binding motifs in common with NuoG, but with divergent carboxy termini. A NuoE homologue was found as a 24-kDa protein with a putative hydrogenosomal targeting pre-sequence¹¹ and termed Ndh24. This sequence is 41% identical to the mitochondrial complex I 24-kDa subunit, and 36% identical to proteobacterial NuoE. *T. vaginalis* Ndh24 has conserved cysteine residues (Fig. 1c) shown to be involved in [2Fe-2S] cluster ligation in *Paracoccus denitrificans* NuoE¹⁴.

To investigate interactions between Ndh51 and Ndh24, we created *T. vaginalis* transformants expressing C-terminal haemagglutinin (HA)-tagged¹¹ Ndh51 and Ndh24. Hydrogenosomes from these transformants were subjected to blue native polyacrylamide gel electrophoresis (BN-PAGE) to determine whether Ndh51 and Ndh24 form stable complexes. Western analyses revealed that tagged proteins Ndh51HA₂ and Ndh24HA₂ from both transformants migrated at about 90 kDa (Fig. 2a), indicating an interaction

between the two proteins. To confirm this, hydrogenosomes were solubilized under non-denaturing conditions and subjected to co-immunoprecipitation with anti-HA antibodies (Fig. 2b); proteins were identified by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Ndh51HA₂ co-immunoprecipitated Ndh24, and Ndh24HA₂ co-immunoprecipitated Ndh51 (Fig. 2b, lanes 1–3), confirming that Ndh24 and Ndh51 form a stable complex. A similar heterodimeric flavoprotein complex (FP) between *P. denitrificans* NuoF and NuoE forms a catalytically active NADH dehydrogenase¹⁵.

To probe for protein partners of the *T. vaginalis* FP complex, we treated hydrogenosomes with the membrane-permeable cleavable crosslinker dithiobis(succinimidylpropionate) (DSP) before co-immunoprecipitation. In addition to recovering the crosslinked, lower-abundance FP complex from Ndh51HA₂ and Ndh24HA₂ hydrogenosomes (Fig. 2b, lanes 4–6), a 60-kDa protein was identified as malic enzyme (MAE) subunits A (ref. 16), B (ref. 16), G and H (Supplementary Fig. 1) and a 120-kDa protein was identified as PFOR, subunit A (ref. 17). Both are crucial enzymes in the central hydrogenosomal carbohydrate metabolism pathway¹ (Fig. 3b). Heterotetrameric MAE catalyses the oxidative decarboxylation of malate to pyruvate and preferentially uses NAD⁺ as co-factor^{16,18}. Our results indicate that both Ndh24 and Ndh51 are in close proximity to the MAE heterotetramer, because four distinct proteins were recovered (Supplementary Fig. 1). PFOR, which forms a

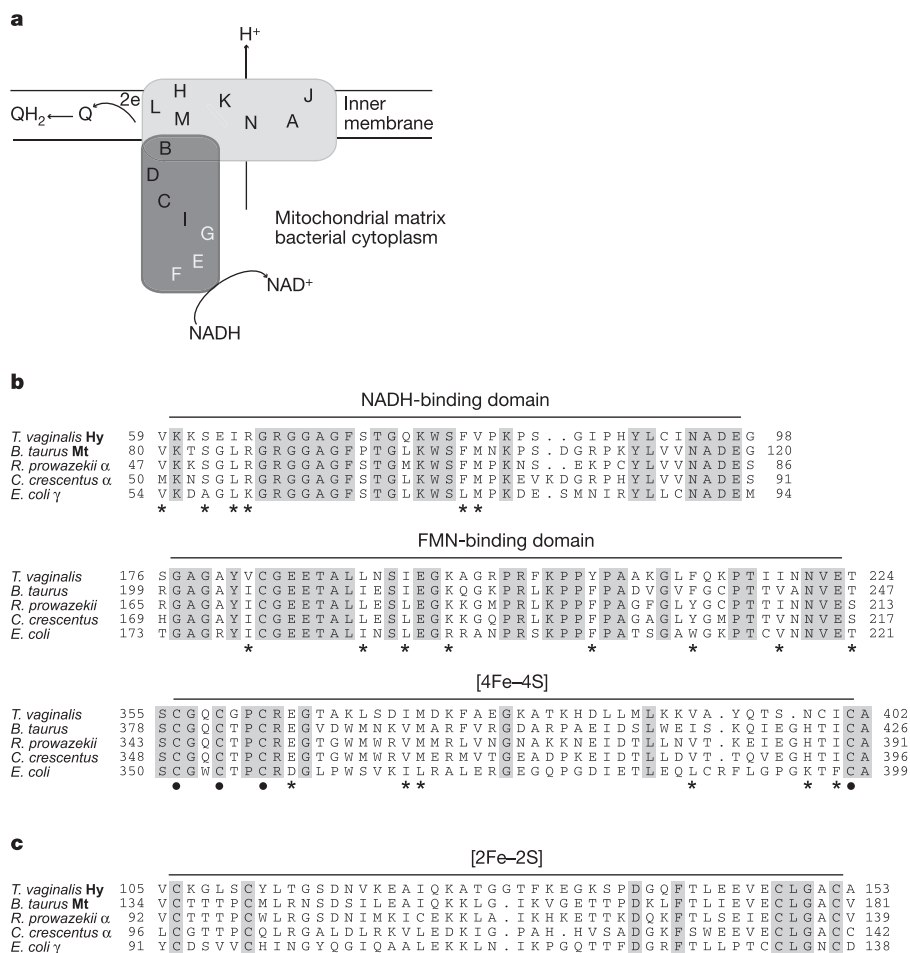


Figure 1 Complex I proteins. **a**, Subunits E, F and G form the NADH dehydrogenase module in proteobacteria and mitochondria. NuoB/NuoD may be involved in quinone binding, and membrane-integrated subunits are possibly involved in quinone binding or proton translocation¹². Light and dark grey boxes indicate membrane-integrated and peripheral subunits respectively. Q, quinone; QH₂, quinol. **b**, Alignment of NuoF functional domains¹³. Hy, hydrogenosomal; Mt, mitochondrial; α and γ represent α- and γ-proteobacteria respectively. Black circles indicate Fe-S cluster-ligating cysteine residues; asterisks and shaded boxes show similar and identical residues respectively. **c**, Alignment of the Fe-S binding domain of NuoE homologues.

membrane-bound dimer, decarboxylates pyruvate to generate acetyl-CoA¹⁷. The CoA moiety from acetyl-CoA is transferred to succinate to form succinyl-CoA, and ATP is produced by substrate-level phosphorylation catalysed by succinate thiokinase. Electrons generated by PFOR are transferred to a [2Fe-2S]-type ferredoxin (Fd) that is reoxidized by an [Fe]-only hydrogenase¹ (Fig. 3b).

We define the supercomplex containing FP, MAE and PFOR as the hydrogenosomal oxidoreductase complex (HOC). To determine whether HOC transfers electrons to quinones, we examined the quinone content of hydrogenosomal lipid extracts (Fig. 3a) using an electrochemical detector (ECD) connected to a high-performance liquid chromatography (HPLC) system. This system provides a detection minimum of 0.3 pmol. No quinone was recovered from hydrogenosomes (Fig. 3a, bottom panel), although internal standards consisting of mitochondrial-type or bacterial-type quinones¹⁹ were recovered (Fig. 3a, top and middle panels). Moreover, we found no homologues to quinone biosynthesis genes²⁰ in *T. vaginalis* genome databases. It therefore seems that HOC functions independently of quinones.

We propose that the Ndh51/Ndh24 FP subcomplex in HOC has a similar function to complex I FP in oxidizing NADH to NAD⁺, with the concomitant transfer of electrons to a terminal [Fe]-type hydrogenase³ related to their ubiquitous partner NuoG (Fig. 3b). Interactions of complex I-like proteins with NAD⁺-reducing hydrogenases in a quinone-independent manner are well documented in *Ralstonia eutropha*, which oxidizes hydrogen and reduces NAD⁺, in a reverse reaction to that proposed here²¹. We have previously shown that hydrogen production was decreased by only 50% in Fd-depleted hydrogenosomes despite drastically lower PFOR activity or concentration²². MAE activity and NADH oxidation were unaffected in those hydrogenosomes²², indicating the absence of coupling of these reactions to PFOR/Fd and the existence of a PFOR/Fd-independent hydrogenase (Fig. 3b, HX2).

Complex I genes are arranged in one operon in most α -proteobacteria²³. Up to 12 of these genes—with the notable exception of *nuoe* and *nuof*, which are nuclear-encoded—are present in various mitochondrial genomes^{23,24}, with some clustered in the same order as in proteobacteria²⁴. Phylogenetic analyses of mitochondria-encoded subunits indicate an origin of the core subunits from a common ancestor with α -proteobacteria²⁵, probably the endo-

symbiont. As shown in Fig. 4a, all mitochondrial and most α -proteobacterial relationships were recovered in phylogenetic reconstructions of NuoF, as reported previously³. Similar relationships were found for NuoE (Fig. 4b), as expected from an endosymbiotic origin of mitochondrial *nuof* and *nuoe* from an α -proteobacterial-type ancestor as part of a *nuo* operon. In contrast, neither *T. vaginalis* Ndh51 (Fig. 4a) nor Ndh24 (Fig. 4b) groups with mitochondrial homologues. Ndh51 seems to originate from an α -proteobacteria ancestral to and distinct from those closely related to mitochondria (Fig. 4a). We included both isoforms of NuoF and NuoE from

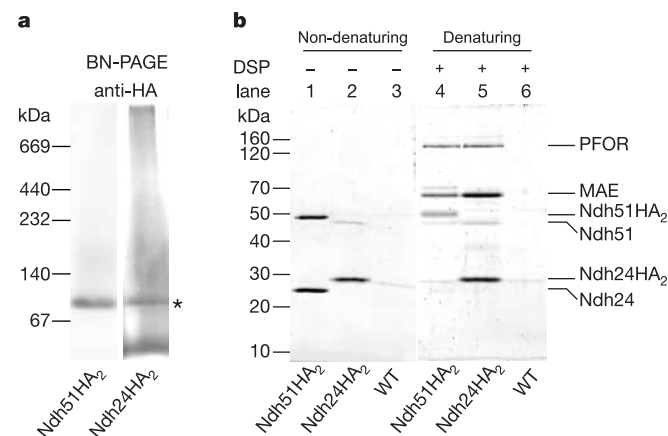


Figure 2 *T. vaginalis* Ndh protein-protein interactions. **a**, Ndh51HA₂ and Ndh24HA₂ each form a stable complex of about 90 kDa (indicated by the asterisk) as shown by BN-PAGE western analyses. **b**, Under non-denaturing solubilization conditions, Ndh24 is associated with Ndh51HA₂ (lane 1) and Ndh51 is associated with Ndh24HA₂ (lane 2). Both Ndh51HA₂ (lane 4) and Ndh24HA₂ (lane 5) can be crosslinked to PFOR and MAE in intact hydrogenosomes and co-immunoprecipitated after lysis under denaturing conditions. Proteins were separated by SDS-PAGE and revealed by Sypro Ruby staining. Bands in the wild-type (WT) lanes correspond to antibody chains.

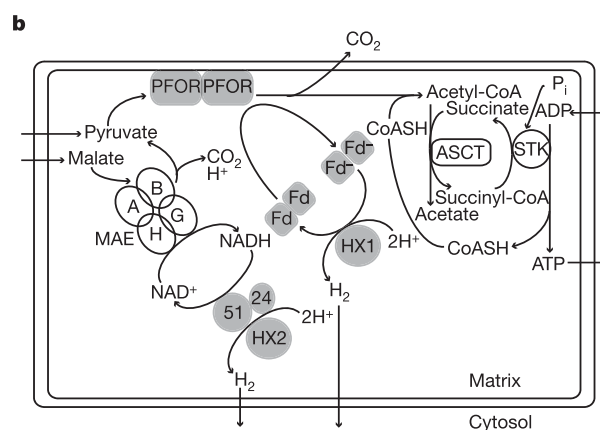
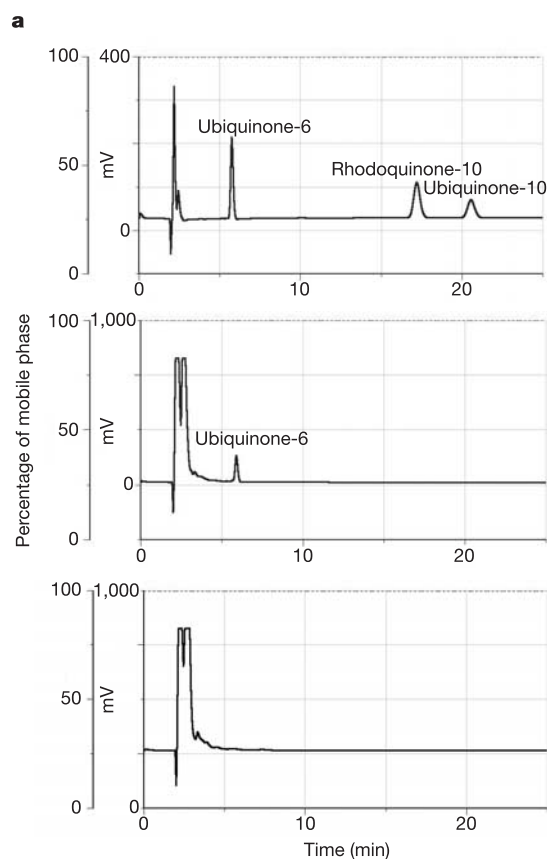


Figure 3 Hydrogenosomal carbohydrate metabolism. **a**, Output from HPLC/ECD system. Quinone standards were used to calibrate the system (top), hydrogenosomal lipid extracts were run with an internal standard (middle), and no quinones were detected in hydrogenosomal extracts alone (bottom). **b**, Partial carbohydrate metabolism pathways. CoASH, coenzyme A; ASCT, acetate:succinate CoA transferase; STK, succinate thiokinase; 51, Ndh51; 24, Ndh24; HX1, H₂-ferredoxin oxidoreductase; HX2, hydrogenase (proposed). Fe-S cluster-containing proteins are shaded in grey. The display does not reflect relative spatial arrangements of proteins.

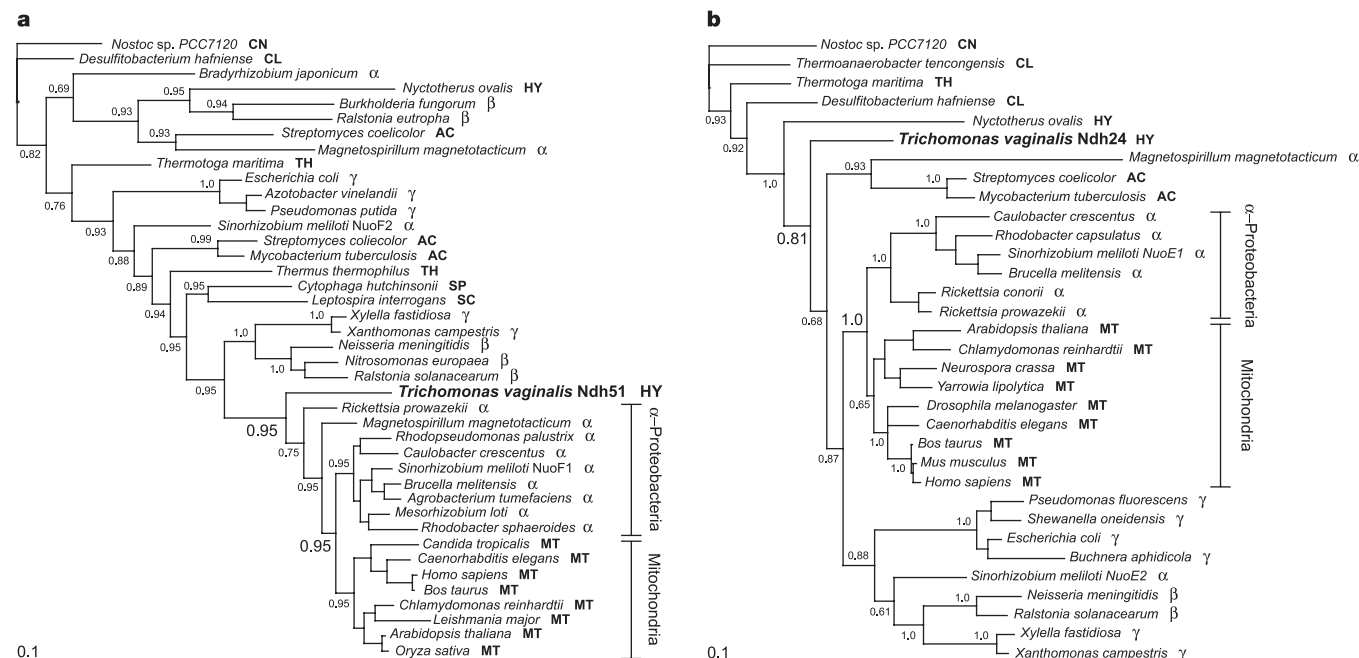


Figure 4 Analyses of NuoF and NuoE phylogenies with MRBAYES³⁰ show that *T. vaginalis* Ndh51 and Ndh24 do not belong to the mitochondrial clade. **a**, Phylogeny of NuoF. Similar topologies for mitochondrial/α-proteobacterial/*T. vaginalis* relationships were recovered with maximum-parsimony and maximum-likelihood methods. **b**, Phylogeny of NuoE. CN,

cyanobacteria; CL, Clostridiales; HY, hydrogenosome; MT, mitochondrion; AC, actinobacteria; TH, Thermotogales; SP, sphingobacteria; SC, spirochaetes; α, β and γ represent α-, β- and γ-proteobacteria respectively. Numbers indicate posterior probabilities of selected branch partitions, where 1.0 indicates maximum support.

α-proteobacteria that have two sets of *nuo* genes, for example *Sinorhizobium meliloti*, and neither set branched with either Ndh51 (Fig. 4a) or Ndh24 (Fig. 4b). Several anaerobic ciliates and chytrids, phylogenetically unrelated to *Trichomonas* but possessing mitochondriate sister groups, contain hydrogen-producing organelles that most probably arose by the conversion of mitochondria through the acquisition of fermentative enzymes. These organelles have many more features in common with mitochondria than *Trichomonas* hydrogenosomes⁴. Such a chytrid, *Nyctotherus ovalis*, bears a fusion of the genes *hydrogenase–nuoe–nuof* with a conceptual hydrogenosomal-type N-terminal presequence²⁶. The relevant Nuo moieties of this fusion protein do not group with either Ndh51 (Fig. 4a) or Ndh24 (Fig. 4b), but the NuoF-like moiety groups with β-proteobacteria as previously observed³, which may indicate its acquisition by horizontal gene transfer. We found no evidence of horizontal gene transfer²⁷ to account for the origin of Ndh51 and Ndh24 in *T. vaginalis*. Furthermore, a long-branch attraction artefact cannot account for the topology of these sequences (Supplementary Fig. 2). Thus, our phylogenetic analyses indicate that NADH dehydrogenase proteins in HOC do not share a common origin with mitochondrial complex I proteins. These data do not support the ‘hydrogen hypothesis’, which stipulates the concurrent creation of a eukaryote with a proto-organelle derived from a single α-proteobacterium that possessed both mitochondrial respiratory complexes and hydrogenosomal fermentative pathways²⁸. Furthermore, enzymes involved in the hydrogenosomal fermentative pathway do not show phylogenies that support this hypothesis^{2,3,18}.

The simplest explanation for these observations is that the *Trichomonas* hydrogenosome does not originate from the proto-mitochondrial endosymbiont. In contrast, chaperonin phylogenies^{8–10} and protein translocation compatibilities are consistent with a common endosymbiotic origin for these two organelles^{7,11}. In this regard, it is notable that plastids and mitochondria have independently recruited similar proteins as building blocks for preprotein translocases in at least three cases⁷. Thus, protein translocation compatibilities between hydrogenosomes and mito-

chondria might be simply a reflection of using similar precursors⁷. Taking present data into account, resolving whether the *Trichomonas* hydrogenosome arose linearly from proto-mitochondria lies in determining which is more parsimonious: to have lost and reacquired two NADH dehydrogenase proteins involved in the central hydrogenosomal metabolic pathway, or to have acquired chaperonins essential for organelle biogenesis from a failed endosymbiotic event^{10,29} distinct from successful independent endosymbioses that led to contemporary mitochondria and trichomonad hydrogenosomes. Further comparative proteomic analyses will be required to resolve this issue. □

Methods

Organism, and hydrogenosomal proteomic analyses

Trichomonas vaginalis strain T1 was used throughout. Hydrogenosomes were isolated and fractionated as described¹¹. Precipitated proteins were solubilized in 2% CHAPS, 2% ASB-14, 7 M urea, 2 M thiourea, 50 mM dithiothreitol and 0.1% ampholytes, and separated by isoelectric focusing followed by 8–16% SDS-PAGE. Proteins were stained with Sypro Ruby (Bio-Rad), excised and digested with trypsin for matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF; Waters-Micromass) or LC-MS/MS (Agilent Technologies). Peptide mass fingerprints were used to search *T. vaginalis* sequence databases using Mascot (Matrix Science). Preliminary sequence data for *ndh51* (contig 870304), *ndh24* (contig 900004), *maeG* (contig 919561) and *maeH* (contigs 857560 and 901330) were obtained from The Institute for Genomic Research (TIGR).

Transformants and protein complex analyses

Protein-coding *ndh51* and *ndh24* sequences were amplified from *T. vaginalis* genomic DNA with primer pairs Ndh51F/Ndh51R and Ndh24F/Ndh24R (Supplementary Table 1) and used to replace the *hmp31* cassette of the pHmp31-(HA)₂ plasmid¹¹ to generate *T. vaginalis* transformation constructs. Transformants were selected that expressed Ndh51HA₂ and Ndh24HA₂, each with a C-terminal dihaemagglutinin tag.

Hydrogenosomes were solubilized (1 mg ml^{−1}) in 1% Triton X-100, 0.3 M NaCl, 10% glycerol, 20 mM MOPS pH 8.0, 1 mM MgCl₂. Proteins were size-separated by 6–16% BN-PAGE, transferred to poly(vinylidene difluoride) membranes, immunodecorated with monoclonal anti-HA antibody (Covance) and detected by enhanced chemiluminescence (Amersham).

For the isolation of complexes, hydrogenosomes were solubilized (1 mg ml^{−1}) in buffer T (1% Triton X-100, 0.3 M NaCl, 0.1 M Tris pH 7.5, 0.5% bovine serum albumin (BSA), 1 mM β-mercaptoethanol). In parallel, intact hydrogenosomes (10 mg ml^{−1}) were crosslinked with 1 mM DSP (Pierce Biotechnology, Inc.), denatured with 1% SDS and diluted 1:9 with buffer T. For co-immunoprecipitation, precleared lysates (1 mg) were

incubated with 20 μ l Sepharose-immobilized monoclonal anti-HA antibodies (Covance). Beads were washed with buffer T without BSA before elution. DSP-induced crosslinks in eluted proteins were thiol-cleaved before separation by 8–16% SDS–PAGE and detection by Sypro Ruby (Bio-Rad).

Quinone analyses

Lipid extractions and quinone detection were performed as described¹⁹. Hydrogenosomes (5.5 mg protein) were extracted and resuspended in 150 μ l 9:1 methanol/ethanol, of which 50 μ l was injected onto an HPLC system linked to an ECD.

Sequence analyses

Accession numbers for sequences used to reconstruct NuoF and NuoE phylogenies are listed in Supplementary Tables 2 and 3. NuoF sequences were aligned with CLUSTALX. NuoE sequences were aligned with Wisconsin Package Version 10.2 programs (Genetics Computer Group). A profile hidden Markov model (HMM) was built from *Escherichia coli*, *Neurospora crassa*, *Bos taurus*, *Paracoccus denitrificans* and *Thermus thermophilus* sequences with HMMBUILD. Additional sequences were aligned to the profile with HMMALIGN. Both alignments were edited to remove C- and N-terminal extensions. Analyses of NuoF and NuoE evolution were performed with MRBAYES³⁰ with the JTT amino-acid substitution model and with two Markov chains Monte Carlo. Chains were run for 100,000 generations, with sampling every 50 generations. The first 5,000 generations were discarded as burn-in. Consensus trees reflecting the more than 50 majority rule were drawn with Treeview, and probabilities of branch partitions were calculated.

Received 25 June; accepted 2 September 2004; doi:10.1038/nature02990.

- Müller, M. The hydrogenosome. *J. Gen. Microbiol.* **139**, 2879–2889 (1993).
- Horner, D. S., Hirt, R. P. & Embley, T. M. A single eubacterial origin of eukaryotic pyruvate:ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. *Mol. Biol. Evol.* **16**, 1280–1291 (1999).
- Horner, D. S., Foster, P. G. & Embley, T. M. Iron hydrogenases and the evolution of anaerobic eukaryotes. *Mol. Biol. Evol.* **17**, 1695–1709 (2000).
- Embley, T. M., van der Giezen, M., Horner, D. S., Dyal, P. L. & Foster, P. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Phil. Trans. R. Soc. Lond. B* **358**, 191–201 (2003).
- Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
- Clemens, D. L. & Johnson, P. J. Failure to detect DNA in hydrogenosomes of *Trichomonas vaginalis* by nick translation and immunomicroscopy. *Mol. Biochem. Parasitol.* **106**, 307–313 (2000).
- Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: from endosymbionts to organelles. *Science* **304**, 253–257 (2004).
- Bui, E. T. N., Bradley, P. J. & Johnson, P. J. A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl Acad. Sci. USA* **93**, 9651–9656 (1996).
- Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. Molecular data suggest an early acquisition of the mitochondrial endosymbiont. *Proc. R. Soc. Lond. B* **263**, 1053–1059 (1996).
- Roger, A. J., Clark, C. G. & Doolittle, W. F. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **93**, 14618–14622 (1996).
- Dyall, S. D. *et al.* Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. *Mol. Cell. Biol.* **20**, 2488–2497 (2000).
- Friedrich, T. & Bottcher, B. The gross structure of the respiratory complex I: a Lego System. *Biochim. Biophys. Acta* **1608**, 1–9 (2004).
- Fearnley, J. M. & Walker, J. E. Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim. Biophys. Acta* **1140**, 105–134 (1992).
- Yano, T., Sled, V. D., Ohnishi, T. & Yagi, T. Identification of amino acid residues associated with the [2Fe–2S] cluster of the 25-kDa (NQO2) subunit of the proton-translocating NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *FEBS Lett.* **354**, 160–164 (1994).
- Yano, T., Sled, V. D., Ohnishi, T. & Yagi, T. Expression and characterization of the flavoprotein subcomplex composed of 50-kDa (NQO1) and 25-kDa (NQO2) subunits of the proton-translocating NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *J. Biol. Chem.* **271**, 5907–5913 (1996).
- Hrdy, I. & Müller, M. Primary structure of the hydrogenosomal malic enzyme of *Trichomonas vaginalis* and its relationship to homologous enzymes. *J. Eukaryot. Microbiol.* **2**, 593–603 (1995).
- Hrdy, I. & Müller, M. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. *J. Mol. Evol.* **41**, 388–396 (1995).
- Dolezal, P., Vanacova, S., Tachezy, J. & Hrdy, I. Malic enzymes of *Trichomonas vaginalis*: two enzyme families, two distinct origins. *Gene* **329**, 81–92 (2004).
- Jonassen, T., Davis, D. E., Larsen, P. L. & Clarke, C. F. Reproductive fitness and quinone content of *Caenorhabditis elegans* *clk-1* mutants fed coenzyme Q isoforms of varying length. *J. Biol. Chem.* **278**, 51735–51742 (2003).
- Meganathan, R. Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. Lett.* **203**, 131–139 (2001).
- Massanz, C., Schmidt, S. & Friedrich, B. Subforms and *in vitro* reconstitution of the NAD-reducing hydrogenase of *Alcaligenes eutrophus*. *J. Bacteriol.* **180**, 1023–1029 (1998).
- Land, K. M. *et al.* Targeted gene replacement of a ferredoxin gene in *Trichomonas vaginalis* does not lead to metronidazole resistance. *Mol. Microbiol.* **51**, 115–122 (2004).
- Friedrich, T. & Scheide, D. The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett.* **479**, 1–5 (2000).
- Lang, B. F. *et al.* An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**, 493–497 (1997).
- Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
- Akhmanova, A. *et al.* A hydrogenosome with a genome. *Nature* **396**, 527–528 (1998).
- Doolittle, W. F. *et al.* How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Phil. Trans. R. Soc. Lond. B* **358**, 39–57 (2003).

- Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- Doolittle, W. F. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**, 307–311 (1998).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. L. Kerwin for mass spectrometric analyses. Preliminary *T. vaginalis* genome sequence data were obtained from TIGR through the website at <http://www.tigr.org>. Sequencing of the *T. vaginalis* genome was accomplished with support from the NIH. This work was supported by National Institute of Health (NIH) grants (P.J.J. and C.F.C.) and a National Aeronautics and Space Administration Astrobiology grant to UCLA. P.J.J. is a Burroughs Wellcome Scholar in Molecular Parasitology. The UCLA Mass Spectrometry and Proteomics Technology Center (J.A.L.) was established with a grant from the W. M. Keck Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.J.J. (johnsonp@ucla.edu).

The genome of *Cryptosporidium hominis*

Ping Xu^{1,2,3*}, Giovanni Widmer^{4*}, Yingping Wang^{1,2}, Luiz S. Ozaki^{1,2}, Joao M. Alves^{1,2}, Myrna G. Serrano^{1,2}, Daniela Puiu¹, Patricio Manque^{1,2}, Donna Akiyoshi⁴, Aaron J. Mackey^{5,†}, William R. Pearson⁶, Paul H. Dear⁷, Alan T. Bankier⁷, Darrell L. Peterson⁸, Mitchell S. Abrahamsen^{9,10}, Vivek Kapur^{10,11}, Saul Tzipori⁴ & Gregory A. Buck^{1,2}

¹Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23284-2030, USA

²Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, Virginia 23298-0678, USA

³Philips Institute for Oral and Craniofacial Molecular Biology, Virginia Commonwealth University, Richmond, Virginia 23298-0566, USA

⁴Tufts University School of Veterinary Medicine, North Grafton, Massachusetts 01536, USA

⁵Department of Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA

⁶Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22908, USA

⁷MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

⁸Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, Virginia 23298-0614, USA

⁹Department of Veterinary and Biomedical Sciences, University of Minnesota, St Paul, Minnesota 55108, USA

¹⁰Biomedical Genomics Center, University of Minnesota, St Paul, Minnesota 55108, USA

¹¹Department of Microbiology, University of Minnesota, St Paul, Minnesota 55108, USA

* These authors contributed equally to this work

† Present address: Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Cryptosporidium species cause acute gastroenteritis and diarrhoea worldwide. They are members of the Apicomplexa—protozoan pathogens that invade host cells by using a specialized apical complex and are usually transmitted by an invertebrate vector or intermediate host. In contrast to other Apicomplexans, *Cryptosporidium* is transmitted by ingestion of oocysts and completes its life cycle in a single host. No therapy is available, and control focuses on eliminating oocysts in water supplies¹. Two species, *C. hominis* and *C. parvum*, which differ in host range, genotype and pathogenicity, are most relevant to humans^{1–3}. *C. hominis* is restricted to humans, whereas

C. parvum also infects other mammals². Here we describe the eight-chromosome ~9.2-million-base genome of *C. hominis*². The complement of *C. hominis* protein-coding genes shows a striking concordance with the requirements imposed by the environmental niches the parasite inhabits. Energy metabolism is largely from glycolysis. Both aerobic and anaerobic metabolisms are available, the former requiring an alternative electron transport system in a simplified mitochondrion. Biosynthesis capabilities are limited, explaining an extensive array of transporters. Evidence of an apicoplast is absent, but genes associated with apical complex organelles are present. *C. hominis* and *C. parvum* exhibit very similar gene complements, and phenotypic differences between these parasites must be due to subtle sequence divergence.

We generated a ~12-fold sequence and ~8-fold bacterial artificial chromosome (BAC) clone coverage of the genome of *C. hominis* isolate TU502 (ref. 3, Fig. 1, Supplementary Figs 1–8, Supplementary Tables 1 and 2). Alignment of the ~9.2-million-base (Mb) final sequence with the HAPPY map⁴ and chromosomes of the *C. parvum* genome⁵ covered ~9.1 Mb. The eight chromosomes range from ~0.9 to ~1.4 Mb and exhibit 31.7% GC content (compare with 30.3% and 19.4% for *C. parvum* and *P. falciparum*⁶, respectively). The density of 2–50-base-pair (bp) repeats was about 1 per 2,800 bp. The distribution of repeats is biased towards chromosome ends because over 85% are in the telomere-proximal thirds of five of the chromosomes (Supplementary Fig. 9). Two octamers, TGGCGCCA and TGCATGCA, over-represented in other apicomplexans⁴, are

~40-fold and 15-fold over-represented in *C. hominis* (Supplementary Table 3). More than 80% of these are in non-coding sequences, indicating possible regulatory or other conserved function. Forty-five tRNAs, four or five rRNA operons—at least one of each of the two known types (Supplementary Table 4)—and two clusters of three tandem 5S rRNA genes are present. As in *P. falciparum*⁶, two tRNA^{Met} genes are present, suggesting discrete roles in initiation and extension. We estimate that there are ~3,994 genes in *C. hominis*, in comparison with 3,952 genes in *C. parvum* and 5,268 in *P. falciparum*⁶ (Table 1). About 60% exhibit similarity to known genes. The distribution of GO annotations for *Cryptosporidium*, *Plasmodium* and *Saccharomyces*, is remarkably similar (Supplementary Fig. 10), indicating that their phenotypic differences are a reflection of non-conserved or previously unreported gene families of unknown function rather than to the functional specialization of conserved gene families. We estimate that 5–20% of *C. hominis* genes have introns.

Analysis of the *C. hominis* genome shows that the parasite possesses a highly tailored glycolysis-based metabolism, is dependent on the host for nutrients, and is exquisitely adapted for its life cycle (Fig. 2, Supplementary Tables 5 and 6). Glycolysis seems functional, unlike the tricarboxylic acid (TCA) cycle and oxidative phosphorylation. Both an anaerobic pathway using pyruvate:NADP⁺ oxidoreductase (PNO) and an aerobic pathway using an alternative oxidase (AOX) are available for recycling NAD⁺ to NADH. In the former, pyruvate is fermented to acetyl coenzyme A (acetyl-CoA) producing NADPH, which is then reduced to NADP⁺, releasing hydrogen, by a Narf-like [Fe]-hydrogenase, as in *Trichomonas*⁷. Acetyl-CoA is processed by acetate CoA synthase to produce acetate and ATP, as in *Giardia*⁸, yielding four ATP per glucose. Acetyl-CoA can also be processed to ethanol yielding no additional ATP. Under glucose-limited conditions, conversion of acetyl-CoA to acetate, generating two extra ATP per glucose, might be favoured. When glucose is in excess, pyruvate can be converted to lactate or ethanol to regenerate NAD⁺ but no additional ATP. *C. hominis* can also generate ATP by metabolism of glycerol using glycerol-3-phosphate dehydrogenase and triose phosphate isomerase.

C. hominis can convert pyruvate to malate and subsequently to oxaloacetate (OAA), regenerating NAD⁺. However, malate shuttle enzymes—for example, aspartate amino transferase—which process OAA to aspartic acid for export from the mitochondrion, are absent. Cytoplasmic malate could be converted to OAA by a mitochondrial membrane-bound malate dehydrogenase, like the lactate shuttle of *Euglena gracilis*⁹, passing electrons from malate to an electron transport system composed of elements of Complexes I and III and an alternative oxidase system with O₂ as electron acceptor and producing no additional ATP.

Enzymes for metabolism of glycogen, starch and amylopectin are present, which is consistent with suggestions that amylopectin represents an energy reserve for sporozoites¹⁰. Lack of glucose-6-phosphate 1-dehydrogenase and other enzymes of the pentose phosphate pathway suggests that, unlike *P. falciparum* and other apicomplexans⁶, *C. hominis* cannot metabolize five-carbon sugars or nucleotides. Components of β -oxidation, for example enoyl-CoA hydratase and acetyl-CoA C-acyltransferase, are also absent, precluding ATP generation from fatty acids. Enzymes for the catabolism of proteins are also absent.

Major TCA-cycle enzymes—isocitrate dehydrogenase, succinyl-CoA synthase and succinate dehydrogenase—are absent in *C. hominis*. Despite the presence of ubiquinol-cytochrome *c* reductase, NADH dehydrogenase (ubiquinone), H⁺-transporting ATPase and iron-sulphur cluster-like proteins, among others, key components of Complexes II and IV are absent, precluding ATP generation by oxidative phosphorylation. Components of oxidative phosphorylation that are present (parts of Complexes I and III) probably reoxidize NADH in a simplified electron-transport chain,

Table 1 *Cryptosporidium hominis* genome summary

(a) The genome	<i>C. hominis</i>	<i>C. parvum</i>	<i>P. falciparum</i>
Size (Mb)	9.16	9.11	22.85
No. of physical gaps	246	5	93
No. of contigs	1413*	n.a.	n.a.
(G+C) content (%)	31.7	30.3	19.4
Coding regions†			
Coding size (Mb)	6.29	6.80	12.03
Percentage coding	69	74	53
(G + C) content (%)	32.3	31.9	23.7
No. of genes	3,994	3,952	5,268
Mean gene length (bp)	1,576	1,720	2,283
Gene density (bp per gene)	2,293	2,305	4,338
Genes with introns (%)‡	5–20%	5%	54%
Hits nr§	2,331	2,483	n.d.
Percentage hits nr§	58	63	n.d.
Intergenic regions			
Non-coding size (Mb)	2.87	2.32	10.83
Percentage not coding	31	25	47
(G+C) content (%)	30.3	25.6	14.6
No. of intergenic regions	4,003	3,960	6,392
Mean length (bp)	716	585	1,694
RNAs			
No. of tRNA genes	45	45	43
No. of 5S rRNA genes	6	6	3
No. of 5.8S, 18S and 28S	5	5	7
(b) The proteome			
Total predicted proteins	3,994	3,952	5,268
Hypothetical proteins	2,779	2,567	3,208
Gene ontology			
Biological process	1,239	n.d.	1,613
Cellular component	1,265	n.d.	1,586
Molecular function	1,235	n.d.	1,625
Structural features			
Transmembrane domain	786	n.d.	1,631
Signal peptide	421	n.d.	544
Signal anchor	221	n.d.	367

*An additional 673 very short contigs are not assembled and probably include contaminant sequences.

†Excluding introns.

‡Estimated intron content from expressed sequence tags.

§Hits, or putative homologous proteins in the non-redundant protein database.

Hypothetical proteins, proteins without sufficient similarity to any other gene to permit functional assignment; n.a., not applicable; n.d., not determined; physical gaps, those that no existing clone closes; transmembrane domains, TMHMM, Trans Membrane Hidden Markov Model (for prediction of transmembrane helices in proteins); signal peptide and signal anchor, SignalP-2.0. *C. parvum* and *C. hominis* genomes were annotated with identical strategies to permit comparison.

as in some plants and protozoa.

Consistent with previous suggestions¹¹ is the observation that *Cryptosporidium* lacks enzymes for the synthesis of key biochemical building blocks—simple sugars, amino acids and nucleotides. However, starch, amylopectin and fatty acids can be generated from precursors. Interestingly, these *C. hominis* enzymes have minimal similarity to the known biosynthetic enzymes and are potential therapeutic targets.

Enzymes of the TCA, urea and nitrogen cycles and of the

shikimate pathway are absent, indicating that *Cryptosporidium* is an amino-acid auxotroph. The shikimate pathway has been proposed as a potential target for glyphosate-based chemotherapy in other parasites including *Cryptosporidium*. We found no evidence to support this hypothesis. Enzymes that interconvert amino acids are encoded in *C. hominis*, and, unlike *P. falciparum*⁶, *C. hominis* has a large complement of amino acid transporters.

C. hominis lacks enzymes to synthesize bases or nucleosides, but encodes enzymes that convert nucleosides into nucleotides and

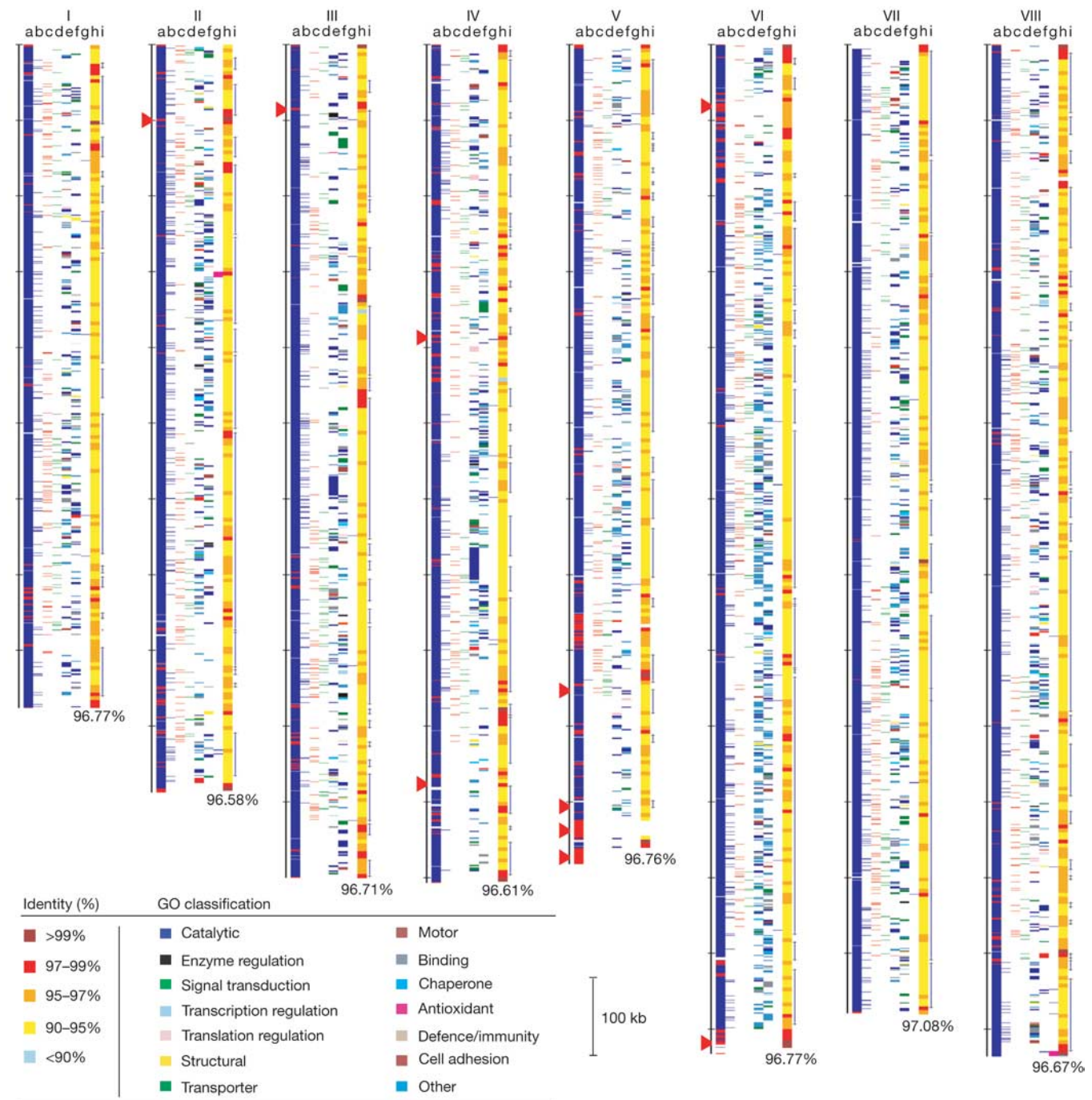


Figure 1 Schematic representation of the *C. hominis* chromosomes. Tracks indicate *C. hominis* contigs (blue), sequence gaps (white) and physical gaps (red) (a); HAPPY⁴ markers (b); positions of the octamers TGGCGCCA (c) and TGCATGCA (d); Gene Ontology (GO) of molecular function for the predicted genes shown by strand (see key, left (e) and right (f)); tRNAs (blue) and rRNAs (magenta) (g); percentage identity to *C. parvum* in 5-kb

windows (see key; average identities are shown at the foot of each chromosome) (h); BAC clone coverage (overlapping clones collapsed to a single line) (i). The scale to the left of each chromosome represents *C. parvum* sequences (red triangles show sequence gaps), with the first base at the top.

interconvert nucleotides. As in other parasites, thymidylate synthase and dihydrofolate reductase of *C. hominis* are encoded as a bifunctional polypeptide, and novel polymorphisms at crucial sites have been proposed to explain *Cryptosporidium*'s resistance to antifolates¹². As previously suggested¹¹, several nucleotide conversion enzymes seem to have a prokaryotic origin.

Fatty-acid biosynthesis in apicomplexans occurs in the apicoplast by means of a type II system including fatty-acid synthase (FAS). However, consistent with the absence of an apicoplast in *Cryptosporidium*¹³ is the observation that *C. hominis* encodes large FAS and polyketide synthase (PKS) enzymes, indicating a type I mechanism. The type I FAS and PKS enzymes of *C. hominis* also have prokaryotic characteristics^{14,15}.

Glycerolipid and phospholipid metabolic pathways for phosphatidylinositol biosynthesis are available in *C. hominis*. 1,2-Diacylglycerol is a precursor for glycosylphosphatidylinositol anchor synthesis. All enzymes required for synthesis of these anchors are apparently present¹⁶.

Polyamines like putrescine, spermine and spermidine are critical for cellular viability, and enzymes required for their synthesis are attractive therapeutic targets. *Cryptosporidium* can synthesize polyamines using arginine decarboxylase rather than ornithine decarboxylase¹⁷. The putative arginine decarboxylase, spermidine synthase and other relevant enzymes encoded by *C. hominis* have diverged significantly from their homologues and are potential therapeutic targets.

C. hominis encodes adenylate cyclase, cyclic-AMP phosphodiesterase and protein kinase A, indicating the presence of the cAMP-mediated signalling pathway (Supplementary Table 7). Trimeric G protein, often involved in the activation of cAMP-mediated signalling, was not found in *C. hominis*, indicating that, as in Kinetoplastida¹⁸ and reminiscent of plants, this pathway is independent of this complex in *C. hominis*. The presence of phosphatidylinositol 3-kinase and phospholipase C indicates that *C. hominis* utilizes phosphatidylinositol phosphate and Ca^{2+} -mediated regulatory mechanisms. The presence of putative Ca^{2+} transporters,

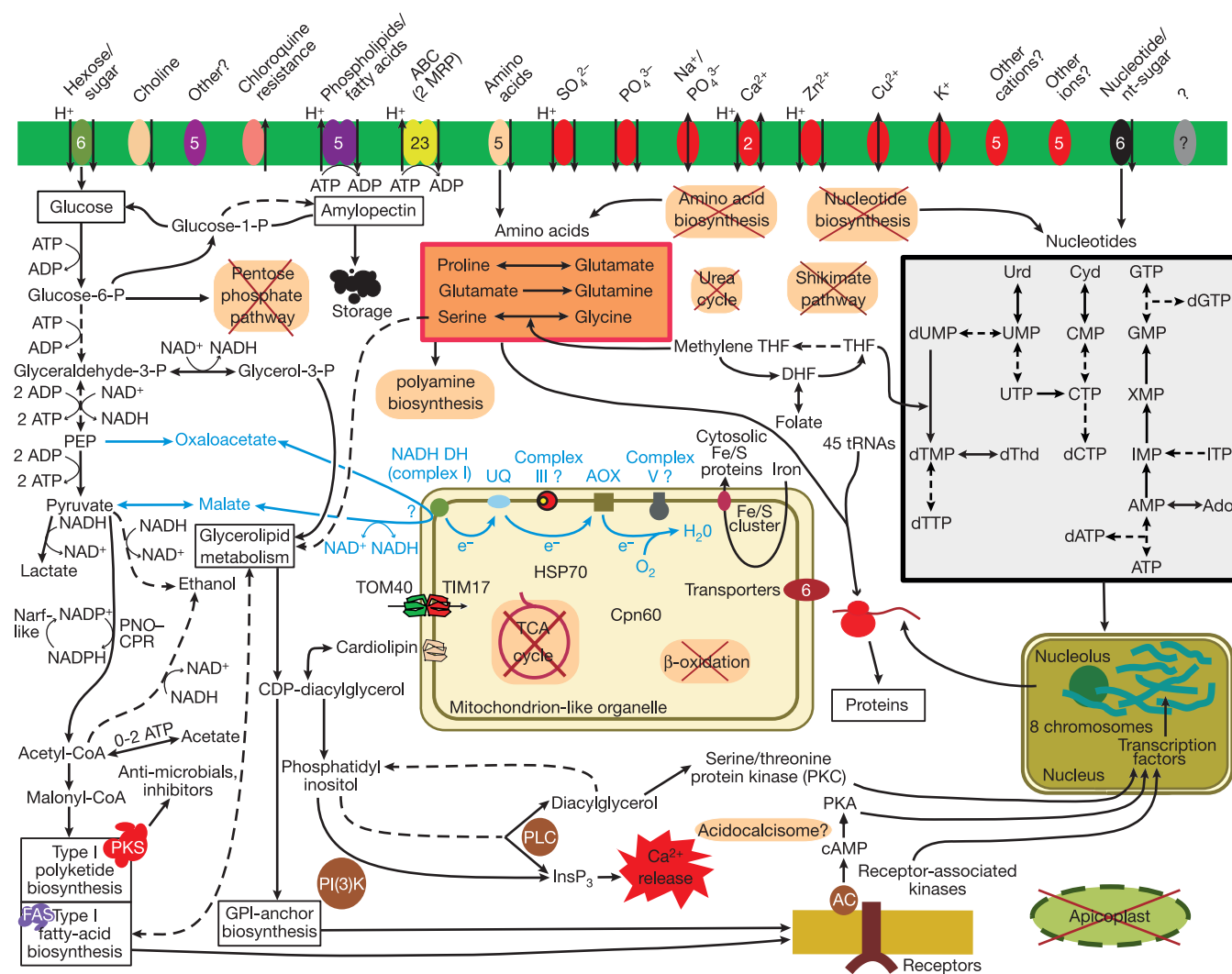


Figure 2 Schematic representation of selective *C. hominis* proteins, enzymes and pathways. The green strip represents the cellular membrane with putative transporters; numbers indicate the number of genes for a given class of transporter. Solid arrows indicate pathways that are present; multistep pathways are indicated with dashed arrows. Components or pathways that are absent are crossed out. Steps or components whose exact nature is questionable are shown with question marks. Blue arrows and names indicate proposed aerobic parts of the metabolism. Abbreviations: ABC, ATP-binding cassette; AC, adenylate cyclase; Ado, adenosine; AOX, alternative oxidase; Cpn60, chaperone 60; Cyd, cytidine; DHF, dihydrofolate; dThd, deoxythymidine; GPI,

glycosylphosphatidylinositol; Hsp70, heat-shock protein 70; InsP₃, inositol phosphate; MRP, multiple-drug-resistance protein; NADH DH, NADH dehydrogenase; Narf-like, nuclear prelamin A recognition factor-like protein; PEP, phosphoenolpyruvate; PI(3)K, phosphatidylinositol 3-kinase; PKA, protein kinase A; PLC, phospholipase C; PKC, protein kinase C; PNO-CPR, pyruvate:NADP⁺ oxidoreductase fused to cytochrome P450 reductase domain; THF, tetrahydrofolate; TIM17, translocase of the inner mitochondrial membrane 17; TOM40, translocase of the outer mitochondrial membrane 40; UQ, ubiquinone; Urd, uridine.

enzymes associated with acidocalcisomes, and calmodulin imply that Ca^{2+} transport and sequestering are functional. Protein kinase C receptors indicate that *C. hominis* has the ability to signal by activation of soluble cytoplasmic receptor-associated kinases.

No mitochondrial DNA sequences were found in *C. hominis*, and the TCA cycle and oxidative phosphorylation are absent (Supplementary Tables 5, 6 and 8). However, a double-membrane-bound organelle generates a proton gradient using cardiolipin and performs some related mitochondrial functions, and mitochondrial marker chaperonin 60 was localized to this structure¹⁹. Core enzymes of [Fe-S] cluster biosynthesis, namely CpFd1, IscU, IscS, mt-HSP70, mtFNR and frataxin, have been reported in *Cryptosporidium*²⁰, and we were not surprised to observe proteins involved in electron transport. We used CDART²¹ to identify [Fe-S] domains in HscB (JAC) and ATM1, which are possibly involved in chaperonin activity of Hsp40/DnaJ type and ABC transport. Thus, *C. hominis*, like the microsporidian *Encephalitozoon cuniculi*²², another obligate intracellular parasite, contains a minimal set of these proteins. These results imply significant mitochondrial function in *C. hominis* and indicate that the previously reported organelle¹⁹ is an atypical mitochondrion.

Cryptosporidium apparently lacks an apicoplast^{13,23}, and searches of the *C. hominis* genome identified no apicoplast-encoded genes (Supplementary Table 9). Some putative nuclear-encoded apicoplast genes, for example acetyl-CoA carboxylase 1 precursor²⁴ and adenyl cyclase²⁵, are present. Others, such as the apicoplast 50S ribosomal protein L33 and the ribosomal L28 and S9 precursor proteins, were not found. The data indicate that *Cryptosporidium* lost an ancestral apicoplast. The presence of D-glucose-6-phosphate ketol-isomerase and 2-phospho-D-glycerate hydrolase, which are similar to plant genes and may be derived from ancient algal endosymbionts, is also indicative that engulfment of the alga that gave rise to the apicoplast preceded the divergence of *Cryptosporidium* from other apicomplexans. One hypothesis is that the acquisition of the type 1 FAS by a progenitor organism obviated the fatty-acid synthesis capabilities of the apicoplast^{14,15}.

The *C. hominis* genome encodes multiple proteins specific for components of the apical complex including micronemes and rhoptries (Supplementary Table 9). No specific dense granule-associated proteins were observed, probably because these proteins diverge rapidly²⁶. However, proteins implicated in the regulation of transport and enhancement of the release of dense granule proteins²⁷ are present. As for *Plasmodium*, a typical Golgi structure is not apparent in *C. hominis*²³. However, the presence of secretory organelles implies the existence of a functional endoplasmic reticulum and Golgi, and *C. hominis* encodes proteins similar to many related components, including the NSF/SNAP/SNARE/Rab machinery, which participates in dense granule release²⁸, and the rhoptry biogenesis mediator activator protein 1, involved in endoplasmic-reticulum-Golgi-organellar protein traffic²⁹. The endoplasmic-reticulum-Golgi-organellar machinery of *C. hominis* therefore seems similar to that of other apicomplexans.

As described above, *C. hominis* exhibits limited biosynthetic capabilities and is apparently dependent on its ability to import essential nutrients such as amino acids, nucleotides and simple sugars. The genome encodes more than 80 genes with strong similarity to known transporters and several hundred genes with transporter-like properties. At least 12 sugar or nucleotide-sugar transporters, five putative amino-acid transporters, three fatty-acid transporters, 23 ABC family transporters including possible multiple-drug-resistance proteins, and several putative mitochondrial transporters are present. Other putative transporters for choline uptake, aminophospholipid transport, ATP/ADP, and others with unclear function, were also identified. These transporters are ideal therapeutic targets (Supplementary Table 10).

Comparison of the genomes of *C. hominis* and *C. parvum* (Fig. 1, Supplementary Table 11) showed that the two genomes are very

similar, exhibiting only 3–5% sequence divergence with no large insertions, deletions (Supplementary Fig. 11) or rearrangements evident. In fact, the gene complements of the two species are essentially identical because the few *C. parvum* genes not found in *C. hominis* are proximal to known sequence gaps (Supplementary Table 1). We therefore conclude that the significant phenotypic differences between these parasites are due to functionally significant polymorphisms in relevant protein-coding genes and to subtle gene regulatory differences.

A striking feature of the *C. hominis* genome is the concordance between its gene complement and the metabolic requirements in the environmental niches of its two primary life-cycle stages—the quiescent oocyst in the nutrient-poor aerobic environment of contaminated water, and the vegetative parasites in the nutrient-rich anaerobic or microaerophilic environment of the host. Oocysts probably persist by aerobically metabolizing stores of complex carbohydrates by means of glycolysis and the alternative electron transport system in the unconventional mitochondrion. Consistent with the lack of the energy-generating TCA cycle, oxidative phosphorylation, β -oxidation and the pentose phosphate pathways is the observation that oocysts are relatively inactive, and the two ATP per glucose from glycolysis can provide sufficient energy. In the host, the parasite can import sugars to fuel glycolysis directly, netting two ATP per hexose. In limiting glucose, an additional two ATP per hexose can be generated either by converting acetyl-CoA to acetate or by means of glycerol metabolism. The residual mitochondrion lacks the TCA cycle and oxidative phosphorylation as expected in an organism that replicates in anaerobic or microaerophilic environments, and a simplified electron transport system for regenerating reducing power is available. Thus, a glycolysis-based metabolism is sufficient to support *Cryptosporidium* in all life-cycle stages.

As previously noted, our analysis shows that *Cryptosporidium* is a mosaic of sequences from diverse progenitors, including the hypothetical endosymbiont alga that formed the apicoplast, the mitochondrion and numerous genes acquired from prokaryotes by lateral transfer. *Cryptosporidium* also exhibits modular gene loss. We assume, on the basis of inference from other apicomplexans and earlier diverging groups such as the Euglenozoa, the Heterolobosea and the Jakobids³⁰, that *Cryptosporidium* progenitors exhibited the TCA cycle, β -oxidation, oxidative phosphorylation, amino acid, nucleotide and sugar biosynthesis, fully competent mitochondria, and a functional apicoplast. Genes associated with these functions are dispersed throughout the genome in *Plasmodium* and, we assume, in the progenitor. However, these systems seem to have been deleted cleanly in *Cryptosporidium*, leaving few identifiable residual genes or pseudogenes. Thus, the *Cryptosporidium* genome is a mosaic resulting from multiple lateral gene transfers and selective gene deletion.

The tailored physiology of *C. hominis* indicates attractive therapeutic targets (Supplementary Table 10), for example: essential transport systems; components of glycolysis; the unique prokaryotic FAS1 and PKS1; starch and amylopectin metabolism; nucleic acid or amino-acid metabolism; the AOX electron transport system; the bifunctional thymidylate synthase-dihydrofolate reductase; and the diverged polyamine synthesis enzymes. Finally, many potential vaccine targets were identified in the *C. hominis* genome (not shown), and, in contrast with other protozoan parasites, no extensive arrays of potentially variant surface proteins were observed, indicating a possible role for immunoprophylaxis for cryptosporidiosis.

The availability of the genome sequence of the human pathogen *C. hominis* is a crucial step forward in our understanding of the biology of this parasite. The gene complement provides very significant insight into its physiology and metabolism, validating previous hypotheses and indicating the possibility of others. New obvious targets for chemotherapy and immunotherapy are already apparent. In short, we expect that the availability of the sequence of

C. hominis will stimulate progress in research on this organism and its pathogenicity, and strategies for intervention in the diseases it causes. □

Methods

A modified whole-genome shotgun strategy was used to sequence the ~9.2-Mb genome of *C. hominis* isolate TU502, which was derived from an infected child from Uganda. DNA was purified from surface-sterilized oocysts, shotgun and BAC clones were constructed, and end sequences were generated. About 220,000 sequence reads from small insert clones, and end sequences from ~2,000 BAC clones averaging ~35 kbp in size, were generated. The data represents a ~12-fold shotgun clone coverage of the genome with a quality score of Phred 20, and a 7–8-fold coverage with BAC clones. The sequences were assembled with Phrap, yielding a ~9.16-Mb assembly, which was structurally and functionally analysed with a variety of available software programs and in-house scripts (see Supplementary Text 1 and 2 for further details and references).

Received 4 May; accepted 6 August 2004; doi:10.1038/nature02977.

1. Tzipori, S. & Ward, H. Cryptosporidiosis: biology, pathogenesis and disease. *Microbes Infect.* **4**, 1047–1058 (2002).
2. Morgan-Ryan, U. M. *et al.* *Cryptosporidium hominis* n. sp. (Apicomplexa: Cryptosporidiidae) from *Homo sapiens*. *J. Eukaryot. Microbiol.* **49**, 433–440 (2002).
3. Akiyoshi, D. E., Feng, X., Buckholt, M. A., Widmer, D. & Tzipori, S. Genetic analysis of *Cryptosporidium parvum*: human genotype 1 isolate passaged through different host species. *Infect. Immun.* **70**, 5670–5675 (2002).
4. Bankier, A. T. *et al.* Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. *Genome Res.* **13**, 1787–1799 (2003).
5. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
6. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
7. Muller, M. in *Evolutionary Relationships among Protozoa* (eds Coombs, G. H., Vickerman, K., Sleight, M. A. & Warren, A.) 109–131 (Kluwer Academic, Dordrecht, 1998).
8. Sanchez, L. B., Galperin, M. Y. & Muller, M. Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming). *J. Biol. Chem.* **275**, 5794–5803 (2000).
9. Jasso-Chavez, R. & Moreno-Sanchez, R. Cytosol-mitochondria transfer of reducing equivalents by a lactate shuttle in heterotrophic *Euglena*. *Eur. J. Biochem.* **270**, 4942–4951 (2003).
10. Petry, F. & Harris, J. R. Ultrastructure, fractionation and biochemical analysis of *Cryptosporidium parvum* sporozoites. *Int. J. Parasitol.* **29**, 1249–1260 (1999).
11. Striepen, B. *et al.* Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl Acad. Sci. USA* **101**, 3154–3159 (2004).
12. Atreya, C. E. & Anderson, K. S. Kinetic characterization of bifunctional thymidylate synthase–dihydrofolate reductase (TS–DHFR) from *Cryptosporidium hominis*: A paradigm shift for TS activity and channeling behavior. *J. Biol. Chem.* **279**, 18314–18322 (2004).
13. Zhu, G., Marchewka, M. J. & Keithly, J. S. *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiol.* **146**, 315–321 (2000).
14. Zhu, G. *et al.* Expression and functional characterization of a giant Type I fatty acid synthase (CpFAS1) gene from *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* **134**, 127–135 (2004).
15. Zhu, G. *et al.* *Cryptosporidium parvum*: the first protist known to encode a putative polyketide synthase. *Gene* **298**, 79–89 (2002).
16. Priest, J. W., Xie, L. T., Arrowood, M. J. & Lammie, P. J. The immunodominant 17-kDa antigen from *Cryptosporidium parvum* is glycosylphosphatidylinositol-anchored. *Mol. Biochem. Parasitol.* **113**, 117–126 (2001).
17. Keithly, J. S. *et al.* Polyamine biosynthesis in *Cryptosporidium parvum* and its implications for chemotherapy. *Mol. Biochem. Parasitol.* **88**, 35–42 (1997).
18. Parsons, M. & Ruben, L. Pathways involved in environmental sensing in trypanosomatids. *Parasitol. Today* **16**, 56–62 (2000).
19. Riordan, C. E., Ault, J. G., Langreth, S. G. & Keithly, J. S. *Cryptosporidium parvum* Cpn60 targets a relict organelle. *Curr. Genet.* **44**, 138–147 (2003).
20. LaGier, M. J., Tachezy, J., Stejskal, F., Kutisova, K. & Keithly, J. S. Mitochondrial-type iron-sulfur cluster biosynthesis genes (IscS and IscU) in the apicomplexan *Cryptosporidium parvum*. *Microbiol.* **149**, 3519–3530 (2003).
21. Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619–1623 (2002).
22. Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
23. Tetley, L., Brown, S. M., McDonald, V. & Coombs, G. H. Ultrastructural analysis of the sporozoite of *Cryptosporidium parvum*. *Microbiol.* **144**, 3249–3255 (1998).
24. Zuther, E., Johnson, J. J., Haselkorn, R., McLeod, R. & Gornicki, P. Growth of *Toxoplasma gondii* is inhibited by aryloxyphenoxypropionate herbicides targeting acetyl-CoA carboxylase. *Proc. Natl Acad. Sci. USA* **96**, 13387–13392 (1999).
25. Muhia, D. K. *et al.* Multiple splice variants encode a novel adenyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **278**, 22014–22022 (2003).
26. Beyer, T. V., Svezhova, N. V., Radchenko, A. I. & Sidorenko, N. V. Parasitophorous vacuole: morphofunctional diversity in different coccidian genera (a short insight into the problem). *Cell Biol. Int.* **26**, 861–871 (2002).
27. Stedman, T. T., Sussmann, A. R. & Joiner, K. A. *Toxoplasma gondii* Rab6 mediates a retrograde pathway for sorting of constitutively secreted proteins to the Golgi complex. *J. Biol. Chem.* **278**, 5433–5443 (2003).
28. Chaturvedi, S. *et al.* Constitutive calcium-independent release of *Toxoplasma gondii* dense granules occurs through the NSF/SNAP/SNARE/Rab machinery. *J. Biol. Chem.* **274**, 2424–2431 (1999).

29. Ngo, H. M. *et al.* AP-1 in *Toxoplasma gondii* mediates biogenesis of the rhoptry secretory organelle from a post-Golgi compartment. *J. Biol. Chem.* **278**, 5343–5352 (2003).
30. Simpson, A. G. *et al.* Evolutionary history of 'early-diverging' eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol. Biol. Evol.* **19**, 1782–1791 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Hendricks, M. R. C. Carvalho, S. Tula, G. Kazanina, J. Power, A. Holzgreffe, N. Ebashi, E. Butt, B. Sutton, S. Millett, W. Vogel and B. Constance for their technical contributions to this project and J. Elhai, D. Mallonee, L. M. Wen, T. Zwierzynski and Z. Chen for their contributions to the planning and performance of this project. This research was supported by grants from the National Institutes of Health.

Author contributions S.T. and G.A.B. contributed equally.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.A.B. (buck@mail2.vcu.edu) or S.T. (saul.tzipori@tufts.edu). The sequences reported in this paper have been deposited in GenBank under the project accession number AAEL000000. Further details of the accession numbers are available in the Supplementary Information.

MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer

Catherine M. Shachaf¹, Andrew M. Kopelman¹, Constadina Arvanitis¹, Åsa Karlsson¹, Shelly Beer¹, Stefanie Mandl², Michael H. Bachmann², Alexander D. Borowsky³, Boris Ruebner³, Robert D. Cardiff³, Qiwei Yang¹, J. Michael Bishop⁴, Christopher H. Contag² & Dean W. Felsher¹

¹Division of Medical Oncology, Departments of Medicine and Pathology,

²Department of Pediatrics, Stanford University, California 94305, USA

³Department of Pathology, University of California Davis Medical Center, Davis, California 95616, USA

⁴G. W. Hooper Foundation, University of California, San Francisco, California 94143, USA

Hepatocellular carcinoma is generally refractory to clinical treatment¹. Here, we report that inactivation of the MYC oncogene is sufficient to induce sustained regression of invasive liver cancers. MYC inactivation resulted *en masse* in tumour cells differentiating into hepatocytes and biliary cells forming bile duct structures, and this was associated with rapid loss of expression of the tumour marker α -fetoprotein, the increase in expression of liver cell markers cytokeratin 8 and carcinoembryonic antigen, and in some cells the liver stem cell marker cytokeratin 19. Using *in vivo* bioluminescence imaging we found that many of these tumour cells remained dormant as long as MYC remain inactivated; however, MYC reactivation immediately restored their neoplastic features. Using array comparative genomic hybridization we confirmed that these dormant liver cells and the restored tumour retained the identical molecular signature and hence were clonally derived from the tumour cells. Our results show how oncogene inactivation may reverse tumorigenesis in the most clinically difficult cancers. Oncogene inactivation uncovers the pluripotent capacity of tumours to differentiate into normal cellular lineages and tissue structures, while retaining their latent potential to become cancerous, and hence existing in a state of tumour dormancy.

Cancer is largely caused by genomic catastrophes that result in the activation of proto-oncogenes and/or inactivation of tumour-suppressor genes². Even brief inactivation of a single oncogene can

be sufficient to induce sustained tumour regression³, indicating that, at least in some cases, oncogene inactivation may result in the permanent loss of a neoplastic phenotype⁴. Liver cancer is one of the most common solid malignancies in the world⁵, with no effective treatment for most of the individuals who succumb to this neoplasm¹. One of the most commonly activated oncogenes associated with the pathogenesis of liver tumours is the MYC oncogene. Animal models have confirmed that overexpression of MYC can induce hepatocellular carcinoma^{6–8}, whereas inhibition of MYC

expression results in a loss of the carcinoma's neoplastic properties⁹. To address the possibility that MYC inactivation may be effective in treating liver cancer, we have developed a conditional transgenic model whereby we can regulate expression of human MYC in murine liver cells.

We used the Tet system¹⁰ to generate transgenic mice that conditionally express the MYC proto-oncogene in liver cells. We crossed TRE-MYC¹¹ mice with a transgenic line, LAP-tTA¹⁰, where the liver activator protein (LAP) promoter drives expression of the

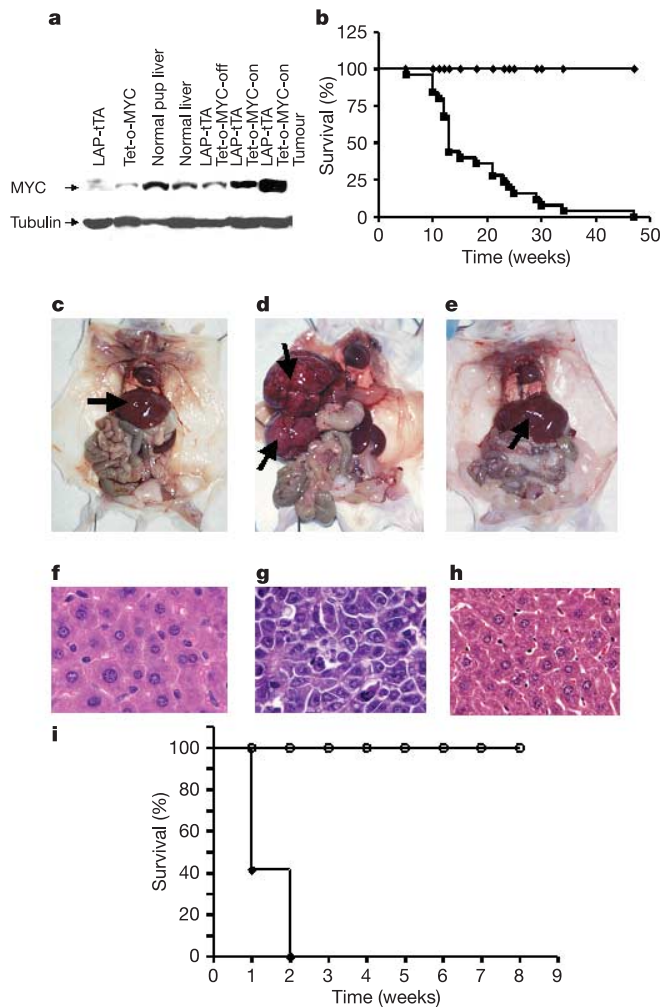


Figure 1 Conditional MYC overexpression in the liver induces hepatocellular cancer whereas MYC inactivation results in sustained tumour regression. **a**, Western analysis for MYC transgene expression. Mice transgenic for both LAP-tTA and tet-o-MYC overexpress MYC in the liver, but mice transgenic for LAP-tTA alone, or where MYC is inactivated, do not. Normal adult and pup livers were used as controls. **b**, Kaplan-Meier survival curve comparing survival of transgenic mice in the presence (squares) or absence (diamonds) of MYC transgene expression. MYC transgene expression was induced in 3-week-old mice by removing doxycycline treatment. Each cohort consists of 25 mice. **c, f**, Mice transgenic for both LAP-tTA and tet-o-MYC and treated with doxycycline exhibited a normal liver grossly (**c**) and histologically (**f**). **d, g**, Mice transgenic for both LAP-tTA and tet-o-MYC that have not been treated with doxycycline succumb to multi-focal liver tumours (**d**) that histologically (**g**) are hepatocellular cancers. **e, h**, Mice with liver tumours that are subsequently treated with doxycycline to suppress the MYC transgene undergo complete regression of their tumours (**e**) and histologically (**h**) the liver appears normal. **i**, Kaplan-Meier survival curve of mice with liver tumours that are either not treated (diamonds) or treated with doxycycline to suppress MYC transgene expression (circles). Each cohort consisted of ten mice.

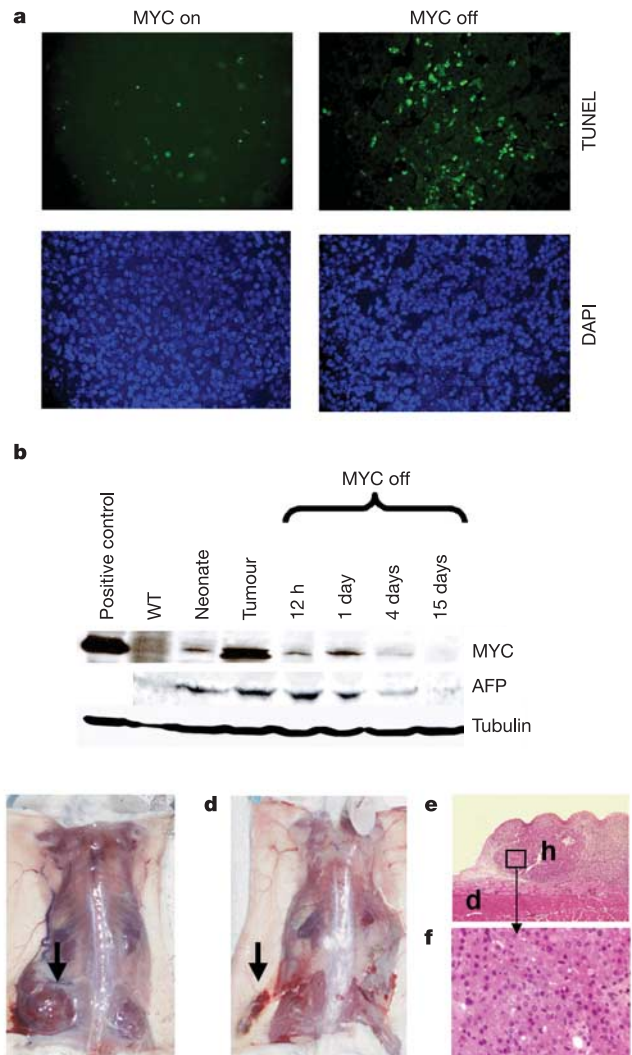


Figure 2 MYC inactivation in liver tumours results in rapid tumour regression associated with loss of expression of tumour markers, differentiation and apoptosis. **a**, TUNEL assay of a liver tumour before and after MYC inactivation. Upper panels show TUNEL staining and lower panels show 4,6-diamidino-2-phenylindole (DAPI) staining of nuclei. Representative data from one of four experiments is shown. MYC inactivation is associated with the differentiation of liver tumour cells into normal hepatocytes. **b**, Western blot analysis for expression of MYC and AFP in normal wild-type (WT) mouse liver, liver of neonatal mice, liver tumour with MYC overexpression and liver tumour where MYC has been inactivated for the indicated periods of time. **c**, Liver tumour cells were transplanted subcutaneously into a SCID mouse (arrow). **d**, MYC inactivation results in tumour regression of the transplanted tumour. **e**, Histological analysis of the tumour site reveals normal-looking hepatocytes (marked as h) within the epidermis (marked as d). **f**, Higher magnification of the differentiated hepatocytes. The experiment was performed five times in six different transgenic tumours in five mice per cohort.

tetracycline trans-activating protein (tTA) in liver cells. Progeny possessing both transgenes expressed MYC, whereas mice with either transgene alone or mice with both transgenes and treated with doxycycline did not express MYC (Fig. 1a). We activated MYC transgene expression in 3-week-old mice by discontinuing doxycycline treatment. Subsequently, all transgenic mice that over-expressed MYC succumbed to liver tumours (Fig. 1b, d, e) with a mean latency of tumour onset of 12 weeks, whereas transgenic mice continuously treated with doxycycline remained free of disease (Fig. 1b, c, f). High levels of MYC expression were detected in tumours but not in liver cells from mice continuously treated with doxycycline, as measured by western analysis and immunohistochemistry (Fig. 1a; see also Supplementary Fig. 1a). Thus in this LAP-tTA/tet-off MYC conditional transgenic mouse model over-expression of MYC in adult mice can reproducibly induce liver cancer. Our results are consistent with previous reports that a MYC transgene can induce liver cancer^{6–8}.

Upon histological analysis the MYC-induced tumours resembled hepatocellular carcinomas and/or hepatoblastomas. The transgenic tumours were locally invasive throughout the liver, frequently

associated with malignant peritoneal effusions that spread via metastasis into the thoracic cavity with invasion into the parenchyma of the lungs. Tumours were readily transplantable into SCID (severe combined immunodeficient) mice, as described below. Liver cancer is particularly refractory to therapeutic intervention^{1,5,12}, therefore, we anticipated that oncogene inactivation in a liver tumour would be even less effective in causing tumour regression than in other types of cancer. Surprisingly, all transgenic mice ($n = 50$) that were moribund with liver tumours exhibited rapid and sustained tumour regression when treated with doxycycline to inactivate MYC transgene expression (Fig. 1i). Upon gross examination 4 weeks after MYC inactivation there was no evidence of tumour persistence, where gross and microscopic liver morphology had been restored to normal (Fig. 1e, h). Thus, the targeted inactivation of MYC alone may effectively induce sustained regression of MYC-induced hepatocellular neoplasia.

To examine the mechanism by which MYC inactivation induces the regression of liver tumours, we treated moribund mice with doxycycline. Within 4 days of MYC inactivation the liver tumours differentiated into normal liver cells accompanied by apoptosis, as

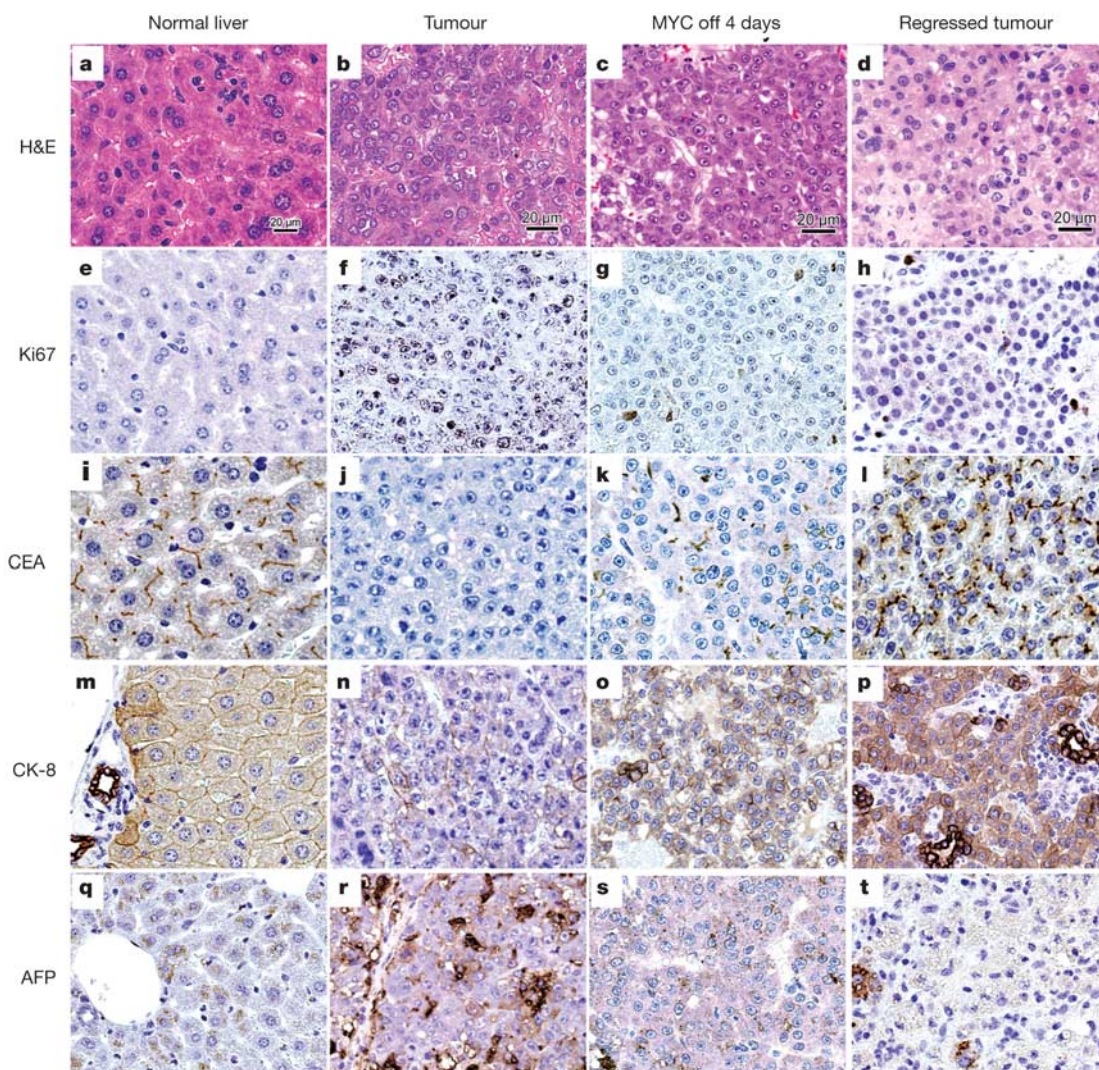


Figure 3 MYC inactivation in liver tumours results in the formation of normal hepatic structures. **a–t**, Normal liver (**a**, **e**, **i**, **m**, **q**), MYC-overexpressing tumour (**b**, **f**, **j**, **n**, **r**), tumour where MYC has been inactivated for 4 days (**c**, **g**, **k**, **o**, **s**) and regressed tumour (**d**, **h**, **l**, **p**, **t**). Serial sections were stained with haematoxylin and eosin (H&E; **a–d**).

Immunohistochemical analysis for markers as indicated: **e–h**, Ki67; **i–l**, CEA; **m–p**, CK-8; and **q–t**, AFP. Representative data are shown from one of three experiments. See also <http://imagearchive.compmed.ucdavis.edu/publications/Shachaf>.

determined by TdT-mediated dUTP nick end labelling (TUNEL) assay (Fig. 2a). Within 2 weeks most of the tumour had grossly regressed, demonstrating that MYC inactivation results in the rapid elimination of most of the tumour cells.

To evaluate unambiguously the fate of tumour cells upon MYC inactivation, we transplanted tumour cell suspensions into SCID mice. After tumours were engrafted into these animals MYC was inactivated. One day after MYC inactivation MYC expression had decreased to 30% of previous levels, and after 4 days it was almost undetectable (Fig. 2b). Similarly, protein expression of the embryonic tumour cell marker characteristic for liver cancer, α -fetoprotein (AFP), was reduced to 50% of previous levels after 1 day and virtually abolished after 4 days. Tumours began to regress within the first 3 days after MYC inactivation and completely regressed

within 30 days, with a residual scar persisting at the site of initial transplantation (Fig. 2d). Notably, histological examination of scar tissue present after MYC inactivation at the site of tumour transplantation in the skin of SCID mice revealed normal liver cells (Fig. 2e, f) resembling hepatic lobules (see also below and Fig. 3). We inferred that MYC inactivation might induce the differentiation of liver cancers into normal liver cells. A trivial possible explanation for our results is that normal liver cells were being transplanted along with cancer cells. This seems to be unlikely as MYC inactivation was observed to induce the differentiation of tumours even when they had been serially transplanted into SCID mice for five passages over a 2-yr time span, beyond the expected duration of proliferative capacity for any normal cells. Tumour cells derived from a distant lung metastasis were also able to differentiate into

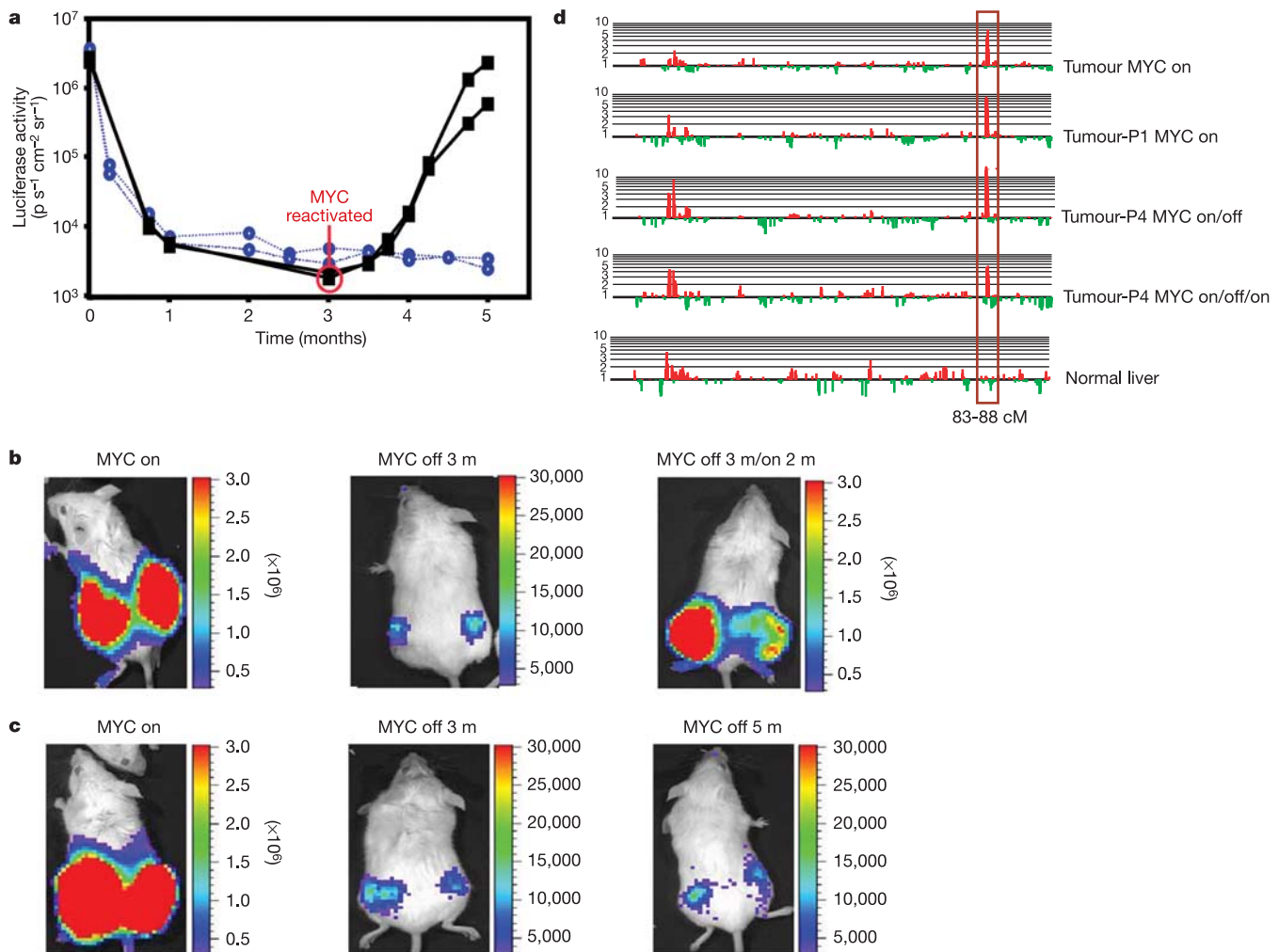


Figure 4 Tumour dormancy observed in liver tumours after MYC inactivation. **a**, Kinetics of tumour regression using *in vivo* bioluminescence imaging of luciferase-labelled liver tumours. Transplanted tumours undergo rapid regression but residual, persisting luciferase activity remains at the site of tumour growth. Upon MYC reactivation tumour growth re-occurred. For visualization of tumour growth, a pseudocolour image representing luciferase light intensity is superimposed over a greyscale reference image of the representative animals in each treatment group: squares, MYC on, then MYC off, and finally MYC on; circles, MYC on then MYC off. Luciferase activity is measured in photons per cm² per second per steradian (p cm⁻² s⁻¹ sr⁻¹). **b**, Representative pictures for mouse where MYC is on (left), MYC is on and then off for 3 months (3 m) (centre), and MYC is on, off for 3 months and then reactivated for 2 months (right). **c**, A representative control mouse is represented for the same time points: MYC on (left), MYC on then off for

3 months (centre), and MYC remains off for 5 months (right). Data are representative of five different experiments with 1–10 animals in each group. **d**, Array CGH analysis of tumours before and after MYC inactivation and reactivation. Array data are shown for a primary tumour tissue (tumour MYC on), transplanted tumour (tumour-P1 MYC on), transplanted tumour (passage 4) where MYC is inactivated (tumour-P4 MYC on/off), transplanted tumour (passage 4) where MYC is reactivated (tumour-P4 MYC on/off/on) and normal liver tissue (normal liver). DNA copy number for chromosome 2 genes is displayed graphically as a ratio (tumour/normal) on a log₁₀ scale along the chromosome. The boxed area marks the region of genomic amplification identified in chromosome 2 shared by this tumour. Red represents gain in gene copy number and green represents deletion in gene copy number.

liver tissues when transplanted into SCID mice. Finally, normal hepatocytes were found to be incapable of persisting beyond 5 days when transplanted into SCID mice (see below). We conclude that upon MYC inactivation, tumours can differentiate into normal liver cells.

To validate further that MYC inactivation induces the differentiation of liver tumour cells, we performed a careful kinetic analysis of the effect of MYC inactivation. Within 4 days most of the tumour cells differentiated into normal hepatocytes and formed normal liver structures (Fig. 3). Tumours lost their neoplastic histological features such as high mitotic index, large nucleoli and hyperchromasia, and they now exhibited a normal nuclear/cytoplasmic ratio. The differentiated tumour cells were Ki67-negative, consistent with their reduced rate of cellular proliferation. Most of the tumour cells lost expression of the immature differentiation marker AFP and instead expressed the hepatocyte and biliary cell liver markers carcinoembryonic antigen (CEA) and cytokeratin 8 (CK-8)¹³, consistent with the formation of sinusoids, bile canaliculi and bile duct cells. Even after prolonged MYC inactivation for up to 8 months some of these differentiated tumour cells persisted, growing under the skin of a SCID mouse (Fig. 3). Moreover, some cells acquired the liver stem cell marker cytokeratin 19 (CK-19)^{13–15} (Supplementary Fig. 2). Thus, upon MYC inactivation most of the liver tumour cells are able to differentiate into hepatocytes and biliary cells, forming bile duct structures.

Next, we examined the consequences of the reactivation of MYC expression. Within 2 weeks of MYC reactivation there was gross evidence for tumour re-growth. Tumours were found to have identical histology to the original transplanted tumour (Supplementary Fig. 1c). These tumours were still dependent on MYC transgene expression, as inactivation of the transgene with doxycycline resulted in tumour regression (Supplementary Fig. 3). Hence, MYC inactivation results in the differentiation of liver tumours, but these tumours retain the latent capacity to regain their neoplastic features upon MYC reactivation.

To examine better the consequences of MYC inactivation and reactivation *in vivo*, we used *in vivo* bioluminescent imaging¹⁶ to visualize tumour growth and response to therapy. We generated liver tumours that were also transgenic for firefly luciferase by crossing the LAP-tTA/tet-o-MYC mice with CMV-GFP-LUC mice¹⁶. The number of tumour cells transplanted and the size of tumour correlated with the light emitted by luciferase activity, allowing us to quantitatively detect as few as 1,000 tumour cells (Supplementary Fig. 4) and to non-invasively examine tumour cell growth and regression in real time.

At 8 months after MYC inactivation luciferase activity was still detectable even when the tumour was not grossly observable (Fig. 4a–c; see also Supplementary Fig. 4). In contrast, normal luciferase-positive liver cells (4×10^7) overexpressing MYC were not detectable 5 days after transplantation into SCID mice. Even when luciferase-positive normal liver cells were transplanted together with luciferase-negative tumour cells, normal hepatocytes were not detectable after 5 days (Supplementary Fig. 5). When MYC expression was restored in the differentiated tumour cells they immediately regained the capacity for proliferation, as demonstrated by an increase in luciferase activity (Fig. 4), and eventually formed grossly visible tumours. Moreover, these tumour cells retained their dependence on MYC expression, as the resumption of doxycycline treatment resulted again in tumour regression (Supplementary Fig. 3). We conclude that MYC inactivation induces a state of tumour dormancy and MYC reactivation is sufficient to restore tumorigenesis.

To confirm further that the liver cells observed upon MYC inactivation and the tumour cells observed upon MYC reactivation were derived from the original tumour cells, array comparative genomic hybridization (CGH) was used to evaluate the presence of genomic signatures. The primary tumour, the serially transplanted

tumour, the regressed tumour after MYC inactivation and the restored tumour after MYC reactivation all possessed the identical regional amplification in a region of chromosome 2 (2H1 83–88 cM) (Fig. 4d). These genetic alterations were unique to this tumour. Two other tumours examined were found to have different changes. Hence, the differentiated cells and the restored tumour are all clonally related to the original tumour.

We have demonstrated that highly invasive and malignant liver cancers exhibit rapid and sustained tumour regression upon MYC inactivation. Loss of MYC expression resulted in the differentiation of tumour cells and eventually most of the cells underwent death. However, tumour cells retained the capacity to differentiate and form normal liver. The reactivation of MYC in these cells restored their neoplastic properties. Hence, MYC inactivation in liver tumours can result in the differentiation of tumour cells into normal liver, but some of these apparently normal cells remain in a state of tumour dormancy.

The diagnosis of invasive liver cancer portends a dismal prognosis and is not amenable to existing therapeutic modalities^{5,12}. Our results suggest that the targeted inactivation of the MYC oncogene may be an effective strategy for the treatment of some liver cancers. In contrast to what we have observed with haematopoietic tumours^{4,11}, only after several serial transplantations did the liver tumours rarely relapse after prolonged MYC inactivation. Relapsed tumours were found to express compensatory, increased levels of L- and N-MYC (Supplementary Fig. 6). It remains to be determined whether MYC inactivation will be effective in the treatment of human liver cancers.

When released from the influence of MYC overexpression, liver tumours *en masse* were able to resume a physiological programme and differentiated into normal hepatic lineages including hepatocytes and biliary cells. Thus, the normalization of MYC expression in liver cancers uncovers their pluripotent capacity to differentiate into normal cellular lineages. Although MYC inactivation resulted in the differentiation and sustained tumour regression of tumours, MYC reactivation was capable of immediately restoring neoplastic properties. Hence, MYC inactivation produces a state of tumour dormancy. Many reports indicate that tumour cells can revert to a state of tumour dormancy^{17–19}. Clinically, it is frequently observed that after therapy tumours exist in a latent state and even after many years are still capable of reverting back to a neoplastic state^{20,21}. Experimentally, tumour dormancy has been induced by means of the suppression of angiogenesis^{22,23} or treatment with anti-idiotypic antibodies¹⁷. We provide here the first report to our knowledge showing that oncogene inactivation can induce tumour dormancy. Liver tumour cells retained the ability to differentiate into multiple hepatic lineages and thus may exist as dormant cancer stem cells^{23–25}. In general, tumour dormancy probably reflects changes in epigenetic regulation associated with the differentiation of tumour cells^{3,26,27}.

The consequences of MYC inactivation apparently depends upon the cellular programming associated with each type of cancer. MYC inactivation in haematopoietic tumours induces differentiation followed by robust apoptosis, which seems to be usually associated with the complete elimination of tumour cells¹¹. Haematopoietic cells seem more poised to undergo apoptosis and renewal from the bone marrow compartment of stem cells as part of their normal physiological programme. MYC inactivation in osteogenic sarcoma is associated with the differentiation of tumour cells into mature bone, but is not associated with apoptosis⁴. Bone cells may be more apt to undergo terminal differentiation without apoptosis. Here, MYC inactivation in liver tumours resulted in the differentiation and eventual death of most of the tumour cells, but some of the tumour cells appeared to have retained the potential to differentiate into multiple liver cellular lineages. However, MYC reactivation was capable of restoring the neoplastic properties to some of these differentiated liver cells, revealing that they existed in a state of

tumour dormancy.

Liver cancers may respond differently from other tumours to oncogene inactivation because the liver has the intrinsic ability to regenerate itself, demonstrating that the liver maintains stem cells. Liver tumours may retain stem cell properties, remaining poised upon MYC inactivation to rapidly differentiate into normal-appearing liver parenchymal cells and duct-like structures. Many recent studies report that cancers frequently consist of cellular subpopulations, some of which have retained stem cell properties and are derived from these cells^{23–25}. MYC seems to be an example of an oncogene that sustains malignant transformation by transforming cells that retain their capacity for cellular differentiation.

Our results suggest a possible model for how MYC activation induces and sustains tumorigenesis in the liver (Supplementary Fig. 7). MYC seems to result in the malignant expansion of immature liver cells with stem cell features, consistent with previous reports suggesting that liver tumours arise from stem cells^{13,14}. Upon MYC inactivation tumour cells began to differentiate and many of them died, but some of the tumour cells showed stem cell properties and differentiated into normal liver. Among these differentiated tumour cells were retained the cancer stem cells, and upon MYC reactivation these cells were a possible source for the re-emergence of the tumour. A less likely possibility for our results is that MYC reactivation resulted in the de-differentiation of the mature hepatocytes that gave rise to tumours. The consequences of oncogene inactivation and reactivation in a given tumour may depend upon the properties of the cellular lineage that has undergone tumorigenesis.

We conclude that there are circumstances when the abatement of oncogene activation is sufficient to resume a normal physiological programme even in cancer cells. Our model system will provide a strategy to interrogate generally how oncogene inactivation uncovers the pluripotent differentiation of cancers and specifically identify the putative liver cancer stem cell. □

Methods

Transgenic mice

Tet-o-MYC transgenic mice have been described previously^{4,11}. LAP-tTA mice¹⁰ were provided by H. Bujard. The generation of transgenic CMV-GFP-LUC mice on a FVB background has been described previously¹⁶.

Tumorigenicity assays

To suppress MYC transgene expression, mice received doxycycline in their drinking water, changed once per week, at a concentration of 100 µg ml⁻¹. For transplantation experiments, tumours were prepared as single cell suspensions (adapted from ref. 28) by incubating liver tumour pieces in HBSS followed by digestion in 1.5 mg ml⁻¹ collagenase in 3 mM KCl, 5 mM NaH₂PO₄, 130 mM NaCl, 10 mM dextrose monohydrate. Cells were washed in PBS twice and re-suspended in PBS, and 10⁷ cells were transplanted subcutaneously into SCID mice.

Histology and immunohistochemistry

Tissues were fixed in 10% buffered formalin, paraffin-embedded and 5-µm paraffin sections were stained with haematoxylin and eosin. Staining was performed using conventional methods fully described in Supplementary Methods.

Western blots

Western analysis was performed using conventional techniques. MYC protein expression was detected using the sc-788 antibody (SantaCruz Biotechnology), AFP (sc-15375, SantaCruz Biotechnology) and α-tubulin (CP06, Oncogene).

In vivo bioluminescence imaging

Transgenic mice were anaesthetized either by injection of a ketamine/xylazine solution (50 µl per 10 g) or by inhalation of an isoflurane/oxygen mixture delivered by the Xenogen XGI-8 5-port gas anaesthesia system. Imaging and image analysis were performed as described previously²⁹ using a cooled CCD camera system (IVIS-100, Xenogen) and LivingImage software (Xenogen).

Array CGH

Genomic DNA labelling, hybridizations and data analysis were performed as described³⁰, with the following modifications: 4 µg of DNA was labelled in a total volume of 50 µl. DNA from tumours were labelled with Cy5 and hybridized to mouse complementary DNA microarrays containing 20,199 different mouse genes (Unigene clusters). Normal FVB/N

mouse spleen DNA labelled with Cy3 was used for reference. For the normal liver sample, normal (FVB/N) male liver DNA was labelled with Cy5 and normal (FVB/N) female DNA was labelled with Cy3 as reference.

Received 8 June; accepted 21 September 2004; doi:10.1038/nature03043.

Published online 10 October 2004.

- Nowak, A. K., Chow, P. K. & Findlay, M. Systemic therapy for advanced hepatocellular carcinoma: a review. *Eur. J. Cancer* **40**, 1474–1484 (2004).
- Bishop, J. M. Molecular themes in oncogenesis. *Cell* **64**, 235–248 (1991).
- Felsher, D. W. Cancer revoked: oncogenes as therapeutic targets. *Nature Rev. Cancer* **3**, 375–380 (2003).
- Jain, M. *et al.* Sustained loss of a neoplastic phenotype by brief inactivation of MYC. *Science* **297**, 102–104 (2002).
- Coleman, W. B. Mechanisms of human hepatocarcinogenesis. *Curr. Mol. Med.* **3**, 573–588 (2003).
- Sandgren, E. P., Quail, C. J., Pinkert, C. A., Palmiter, R. D. & Brinster, R. L. Oncogene-induced liver neoplasia in transgenic mice. *Oncogene* **4**, 715–724 (1989).
- Murakami, H. *et al.* Transgenic mouse model for synergistic effects of nuclear oncogenes and growth factors in tumorigenesis: interaction of c-myc and transforming growth factor alpha in hepatic oncogenesis. *Cancer Res.* **53**, 1719–1723 (1993).
- Wu, Y. *et al.* Recurrent allelic deletions at mouse chromosomes 4 and 14 in Myc-induced liver tumors. *Oncogene* **21**, 1518–1526 (2002).
- Simile, M. M. *et al.* Downregulation of c-myc and Cyclin D1 genes by antisense oligodeoxynucleotides inhibits the expression of E2F1 and *in vitro* growth of HepG2 and Morris 5123 liver cancer cells. *Carcinogenesis* **25**, 333–334 (2003).
- Kistner, A. *et al.* Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 10933–10938 (1996).
- Felsher, D. W. & Bishop, J. M. Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell* **4**, 199–207 (1999).
- Venook, A. P. Hepatocellular carcinoma. *Curr. Treat. Options Oncol.* **1**, 407–415 (2000).
- Thorgeirsson, S. S. Hepatic stem cells in liver regeneration. *FASEB J.* **10**, 1249–1256 (1996).
- Fausto, N. Hepatocyte differentiation and liver progenitor cells. *Curr. Opin. Cell Biol.* **2**, 1036–1042 (1990).
- Fausto, N. & Campbell, J. S. The role of hepatocytes and oval cells in liver regeneration and repopulation. *Mech. Dev.* **120**, 117–130 (2003).
- Hardy, J. *et al.* Bioluminescence imaging of lymphocyte trafficking *in vivo*. *Exp. Hematol.* **29**, 1353–1360 (2001).
- Uhr, J. W., Tucker, T., May, R. D., Siu, H. & Vitetta, E. S. Cancer dormancy: studies of the murine BCL1 lymphoma. *Cancer Res.* **51**, 5045s–5053s (1991).
- Demicheli, R. Tumour dormancy: findings and hypotheses from clinical research on breast cancer. *Semin. Cancer Biol.* **11**, 297–306 (2001).
- Naumov, G. N., MacDonald, I. C., Chambers, A. F. & Groom, A. C. Solitary cancer cells as a possible source of tumour dormancy? *Semin. Cancer Biol.* **11**, 271–276 (2001).
- Uhr, J. W., Scheuermann, R. H., Street, N. E. & Vitetta, E. S. Cancer dormancy: opportunities for new therapeutic approaches. *Nature Med.* **3**, 505–509 (1997).
- Riethmuller, G. & Klein, C. A. Early cancer cell dissemination and late metastatic relapse: clinical reflections and biological approaches to the dormancy problem in patients. *Semin. Cancer Biol.* **11**, 307–311 (2001).
- Holmgren, L., O'Reilly, M. S. & Folkman, J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nature Med.* **1**, 149–153 (1995).
- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
- Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448–452 (2003).
- Al-Hajj, M. B., Wicha, M. W., Weissman, M. & Clarke, I. M. F. Therapeutic implications of cancer stem cells. *Curr. Opin. Genet. Dev.* **14**, 43–47 (2004).
- Laird, P. W. *et al.* Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* **81**, 197–205 (1995).
- Lotem, J. & Sachs, L. Epigenetics wins over genetics: induction of differentiation in tumor cells. *Semin. Cancer Biol.* **12**, 339–346 (2002).
- Leffert, K. S. & Skelly, K. in *Cell Culture Methods for Molecular and Cell Biology* (eds Barnes, D. W., Sirbasku, D. A. & Sato, G. H.) 43–55 (Alan R. Liss, New York, 1984).
- Edinger, M. *et al.* Revealing lymphoma growth and the efficacy of immune cell therapies using *in vivo* bioluminescence imaging. *Blood* **101**, 640–648 (2003).
- Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Bujard for providing us with the LAP-tTA mice; Y. Kim for assistance with the array CGH analysis; L. Germain for the CK-19 antibody; I. Weissman, P. Khavari and J. Sage for critical reading of the manuscript; and N. Bradon, S. Youssef and members of the Felsher laboratory for their suggestions. This work was supported, in part, by the National Cancer Institute (D.W.F. and C.H.C.), the ASCO Young Investigator Award, a Pilot Feasibility Grant from the UCSF Liver Center, a Pilot Award from the Stanford University Digestive Disease Center (D.W.F.), Weiland Family Fellowship (C.M.S.), Howard Hughes Medical Institute Medical Student Fellowship (A.M.K.) and a PHS NRSA training grant (A.K., S.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W.F. (dfelsher@stanford.edu).

Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control

Satoru Mimura*, Takashi Seki*†, Seiji Tanaka† & John F. X. Diffley

Cancer Research UK London Research Institute, Clare Hall Laboratories, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3LD, UK

*These authors contributed equally to this work

† Present addresses: Division of Biological Science, Graduate School of Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan (T.S.); National Institute of Genetics, Division of Microbial Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan (S.T.)

Cyclin-dependent kinases (CDKs) limit the activation of DNA replication origins to once per cell cycle by preventing the assembly of pre-replicative complexes (pre-RCs) during S, G2 and M phases of the cell cycle in the budding yeast *Saccharomyces cerevisiae*^{1,2}. CDKs inhibit each pre-RC component (ORC, Cdc6, Cdt1/Mcm2-7) by different mechanisms. We show here that the mitotic CDK, Clb2/Cdc28, binds tightly to an amino-terminal domain (NTD) of Cdc6, and that Cdc6 in this complex is unable to assemble pre-RCs. We present evidence indicating that this Clb2-dependent mechanism contributes to preventing re-replication *in vivo*. CDK interaction with the NTD of Cdc6 is

mediated by the cyclin subunit Clb2, and could be reconstituted with recombinant Clb2 protein and synthetic NTD peptides. Tight Clb2 binding occurred only when the NTD was phosphorylated on CDK consensus sites. Human CDKs containing cyclins A, B and E also bound specifically to phospho-NTD peptides. We propose that direct binding of cyclins to phosphopeptide motifs may be a widespread phenomenon contributing to the targeting of CDKs to substrates.

Extracts from G1- but not G2/M-arrested cells are able to load the pre-RC components Cdc6 and Mcm2-7 onto yeast DNA replication origin (ARS1)-containing paramagnetic beads³. We asked whether the high CDK activity in G2/M extracts is responsible for preventing pre-replicative complex assembly. Cdc28 (budding yeast homologue of Cdk1) was depleted from a G2/M extract using p13^{suc1} beads^{4,5} (Fig. 1a). After depletion, recombinant Cdc6 (rCdc6) was added to the extract, and pre-RC formation was examined (Fig. 1b). In a mock-depleted extract, ORC bound specifically to the wild-type origin-containing DNA but Cdc6 and Mcm2 loading did not occur. In the Cdc28-depleted extract the loading of both Cdc6 and Mcm2 to the wild-type origin was restored. Thus, CDKs inhibit pre-RC assembly in the G2/M extract.

We next examined pre-RC assembly in extracts from G2/M-arrested cells expressing either wild-type Cdc6 (Cdc6(WT)) or a mutant form of Cdc6 in which the serine or threonine in each of the eight potential CDK phosphorylation sites was mutated to alanine (Cdc6(1-8)A). Figure 1c shows that, in contrast with Cdc6(WT),

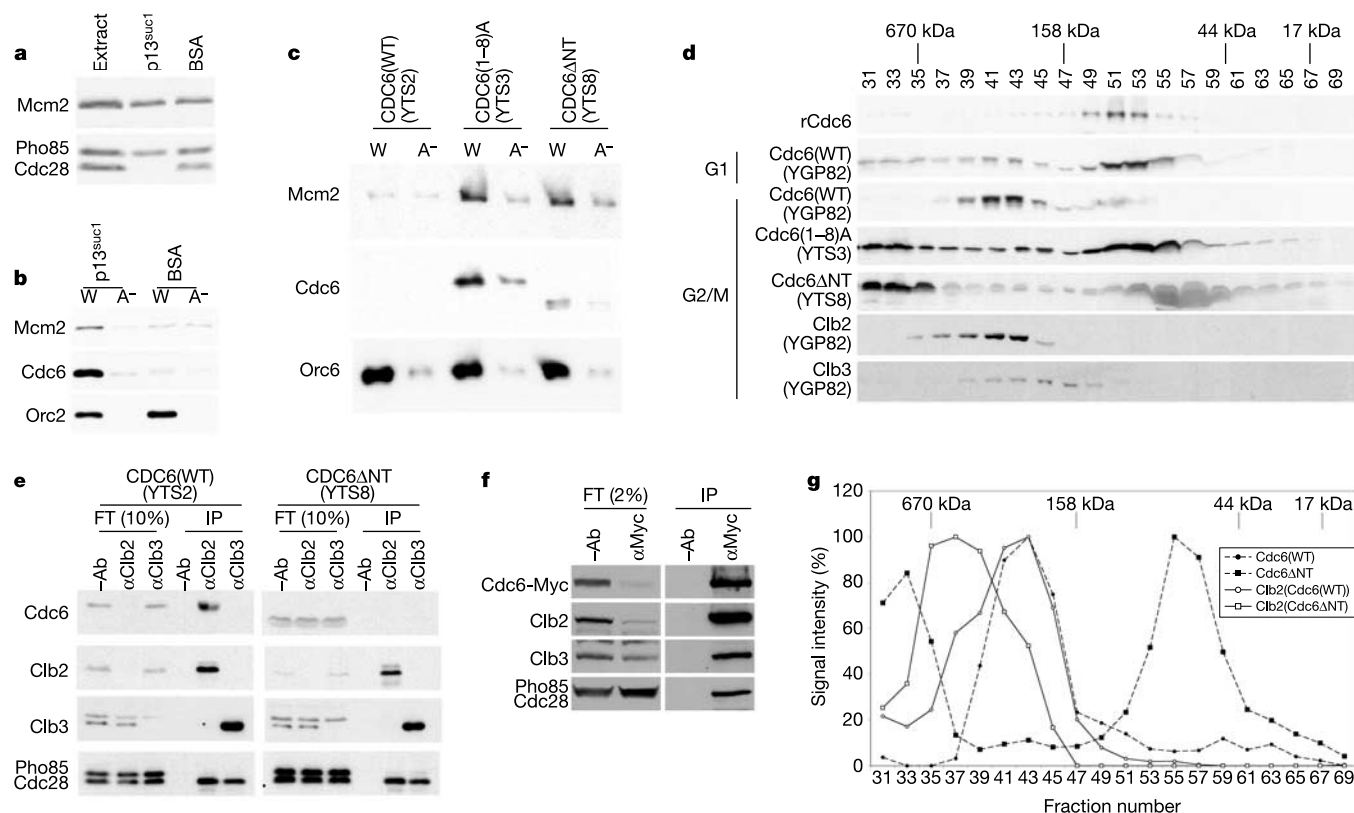


Figure 1 Clb2/Cdc28 inhibition of Cdc6 loading. **a**, G2/M extracts lacking Cdc6 (YGP82, glucose) and depleted with either p13^{suc1} beads or BSA beads were analysed by immunoblotting. **b**, Recombinant Cdc6 (rCdc6) was added to the extracts from **a** and pre-RC assembly was examined. W, wild-type ARS1 beads; A⁻, ARS1 lacking domain A³. **c**, Gal extracts were prepared from cells arrested at G2/M, and pre-RC assembly was examined. **d**, G1 or G2/M Gal extracts were fractionated by gel filtration and analysed by immunoblotting. Molecular mass marker protein positions are indicated above fraction

numbers. **e**, G2/M Gal extracts were immunoprecipitated with the antibodies indicated at the top. Precipitated proteins in **e** and **f** were detected by immunoblotting with antibodies indicated on the left (IP, beads bound; FT, flow through). **f**, G2/M Gal extracts from YLD71 (GAL-CDC6MYC13; a 1.5 h induction with galactose) were immunoprecipitated with anti-Myc antibody (9E11). **g**, G2/M Gal extracts from YTS2 (GAL-CDC6(WT)) or YTS8 (GAL-CDC6ΔNT) were treated as in **d**. Signals were quantified using Adobe Photoshop software.

Cdc6(1–8)A was itself loaded and could load Mcm2 onto ARS1 beads in G2/M extracts. An NTD comprising the first 47 amino acids of Cdc6 contains four potential CDK phosphorylation sites. Figure 1c shows that Cdc6 Δ NT, which lacks the NTD, bound to wild-type ARS1 beads and loaded Mcm2 onto the beads, similar to Cdc6(1–8)A.

Because the NTD is required for CDK-dependent inhibition of Cdc6 but is not required for Cdc6 function in replication⁶, and because CDK phosphorylation of Cdc6 *in vitro* was not sufficient to inhibit Cdc6 function (Supplementary Fig. 1), we considered that the NTD may act by binding a Cdc6 inhibitor. To begin to address this possibility, Cdc6 expressed from the *GAL1,10* promoter

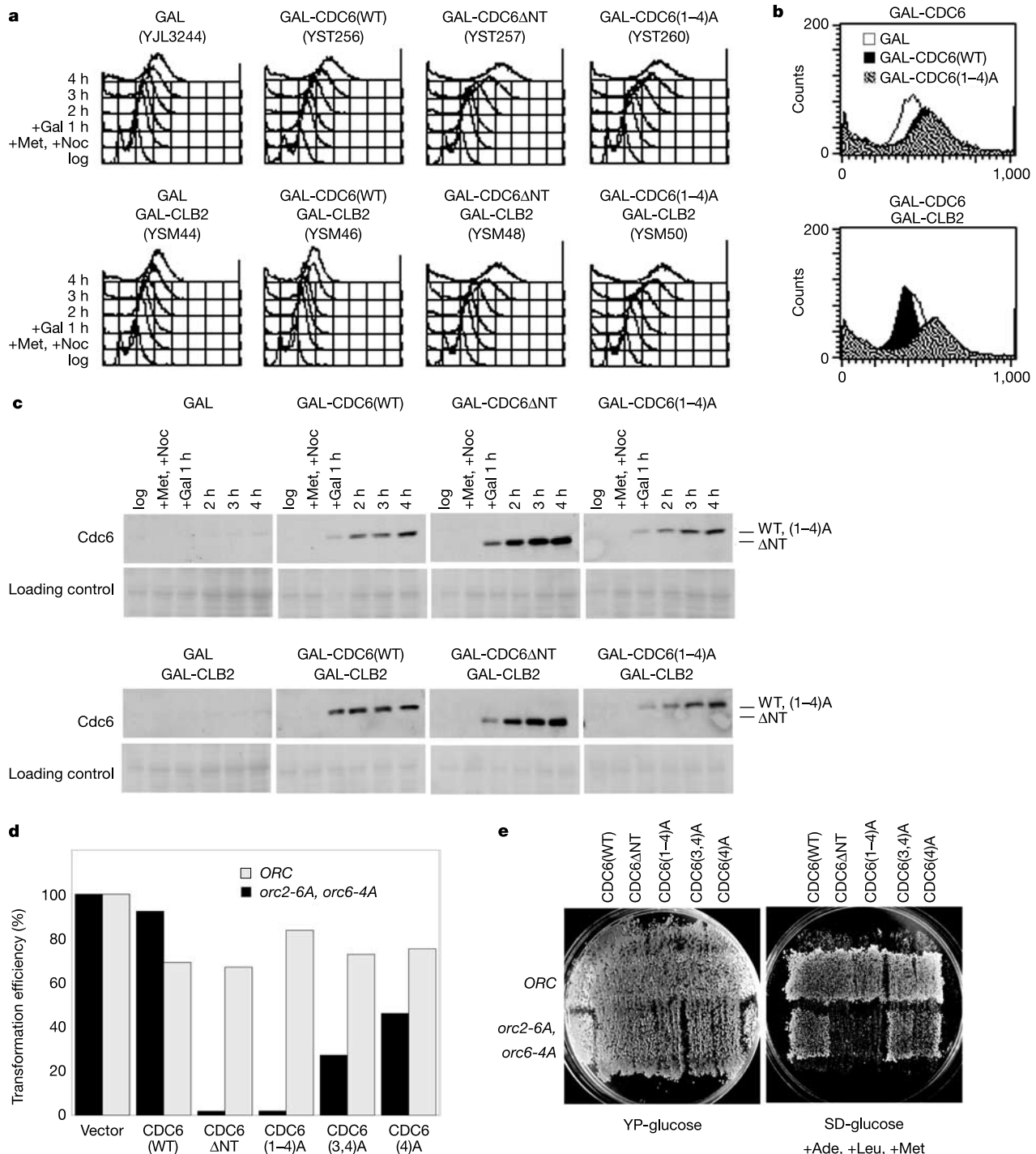


Figure 2 CDK consensus sites in the NTD prevent re-replication *in vivo*. **a**, *MCM7-NLS*, *orc2-6A*, *orc6-4A* mutant strains were arrested in G2/M and the indicated proteins were induced with galactose for the indicated times. Noc, nocodazole. **b**, Merged flow cytometry patterns from **a**, 4 h after galactose addition. Units for x axis are arbitrary units. **c**, Cdc6 proteins from **a** were detected by immunoblotting. **d**, YGP84 (*ORC*⁺; grey boxes) and YJL1737 (*orc2-6A*, *orc6-4A*; black boxes) were transformed with plasmids

expressing the indicated Cdc6 proteins from the *CDC6* promoter. Colony numbers were counted after 3 days. **e**, *MAT α* strains (A364a and YJL1737) and *MAT α* strains (YSM58, YSM59, YSM60, YSM77 and YSM78) were plated after crossing on YPD medium then replica-plated on medium selecting for diploids (SD-glucose + Ade, + Leu, + Met) for 3 days at 30 °C.

(Gal-expressed) was analysed after gel filtration of G1 or G2/M extracts (Fig. 1d). In the G1 extract, Cdc6(WT) was detected in a broad range of fractions, but predominately in fractions of low molecular mass (fractions 49–55), similar to purified rCdc6. In G2/M extracts, however, Cdc6(WT) was predominately found in higher molecular mass fractions (fractions 39–45). In contrast, significant amounts of both Cdc6(1–8)A and Cdc6ΔNT in G2/M extracts were found in lower molecular mass fractions, similar to Cdc6(WT) in G1 extracts. Cdc6 proteins were also present in aggregates of very high molecular mass, caused by overexpression. Taken together, these results are consistent with the possibility that Cdc6(WT) interacts via the NTD with an inhibitor protein(s) during the G2/M phase of the cell cycle, and this interaction requires CDK phosphorylation sites.

Clb2/Cdc28 has previously been shown to interact with Cdc6 by means of the NTD^{7–10}. Figure 1e shows that both Cdc6(WT) and Cdc28 were co-immunoprecipitated from an extract at G2/M with anti-Clb2 antibody (IP). Notably, Cdc6(WT) was almost completely depleted from the extract (flow through; FT) with the Clb2 antibody, indicating that most or all of the Cdc6 protein in these extracts is bound to Clb2. No Cdc6 was detected in Clb3 immunoprecipitates. Other CDK substrates such as Orc6 were not detected in the Clb2 immunoprecipitates (data not shown), suggesting that being a CDK substrate is not sufficient for strong interactions with Clb2. Figure 1e shows that Cdc6ΔNT was not co-immunoprecipitated from the G2/M extract with anti-Clb2 antibody, consistent with the fact that the NTD is required for CDK interaction^{7–9,11,12}. Figure 1f shows that immunoprecipitation of Gal-expressed, Myc-tagged Cdc6 co-immunoprecipitated both Clb2 and Clb3.

This depleted most of Clb2 and a fraction of Clb3 from the extracts. Myc-tagged Cdc6 is more stable than untagged Cdc6 and consequently is expressed at significantly higher levels (data not shown). Thus, Clb2/Cdc28 forms a stable complex with Cdc6 via the NTD, and Cdc6 in this complex is inactive for pre-RC assembly. At higher levels of expression, Cdc6 can also interact with Clb3.

Clb2, but not Clb3, co-fractionated with Cdc6(WT) after gel filtration (Fig. 1d, g), consistent with these proteins forming a stable complex. However, when Cdc6ΔNT was expressed (Fig. 1g) or in the absence of Cdc6 expression (data not shown) in cells at G2/M, Clb2 was predominately found in fractions of higher molecular mass. This indicates that Clb2 normally interacts with other proteins and that Cdc6(WT) can sequester Clb2 from these proteins.

Overexpression of Cdc6ΔNT can induce significant amounts of re-replication in nocodazole-arrested cells when CDK regulation of Mcm2-7 is disrupted by adding a constitutively active nuclear localization sequence (NLS) to Mcm7 and all of the CDK consensus sites in Orc2 and Orc6 are mutated (*orc2-6A, orc6-4A*)¹³. Figure 2a, b shows that expression of either Cdc6ΔNT or Cdc6(1–4)A induced significantly more re-replication in this background than Cdc6(WT). The re-replication induced by Cdc6(WT) was completely suppressed by co-expression of Clb2, whereas the re-replication induced by either Cdc6ΔNT or Cdc6(1–4)A was unaffected by co-expression of Clb2. Therefore, Clb2 can inhibit Cdc6 *in vivo* through CDK sites in the NTD of Cdc6. Figure 2c shows that the levels of Cdc6(1–4)A were essentially the same as Cdc6(WT), consistent with the fact that phosphorylation of the NTD makes no contribution to

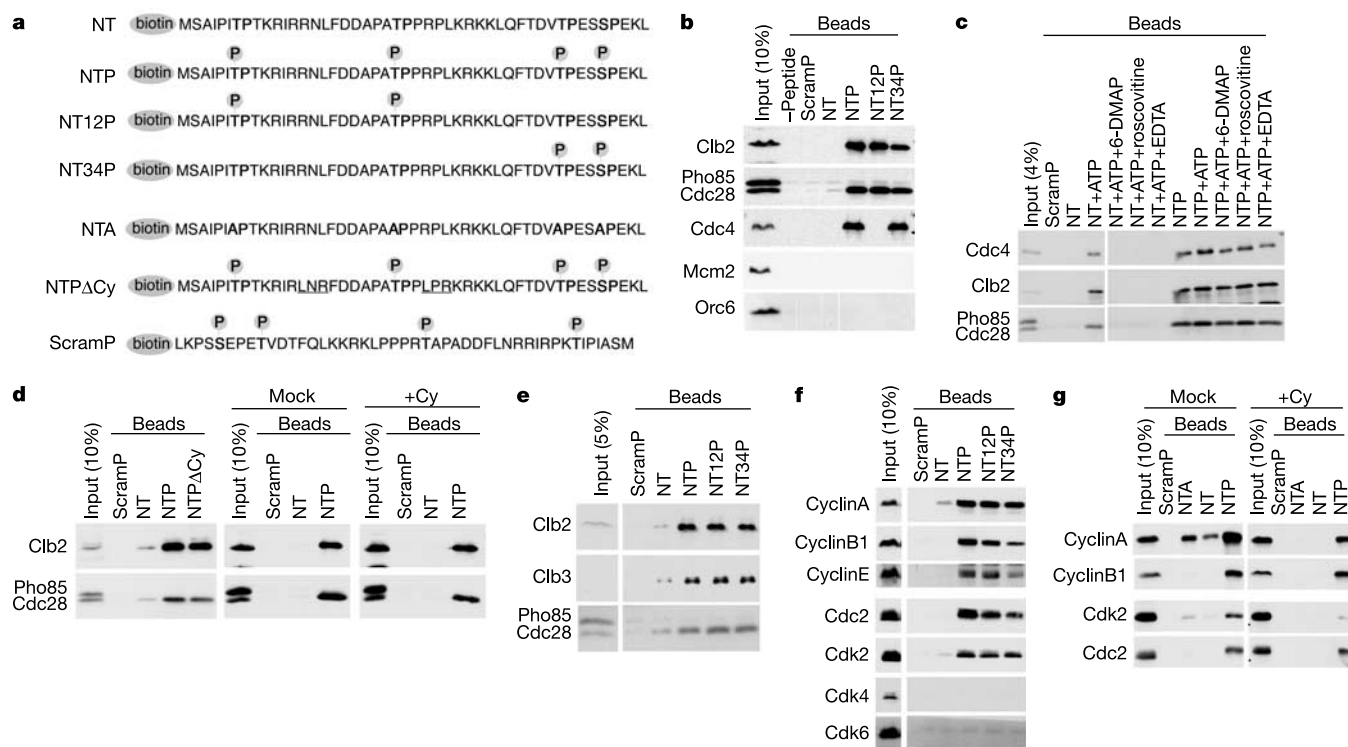


Figure 3 Phosphorylation-dependent binding of CDKs to Cdc6 N-terminal peptides.

a, Peptides contain the first 47 amino acids of Cdc6(WT) and were biotinylated on their N termini. P indicates a phospho-amino acid. **b**, Biotinylated peptides bound to streptavidin-conjugated magnetic beads were incubated (20 min at 24 °C) in G2/M Gal extracts from YGP82 (GAL-CDC6(WT)) in the presence of 3 mM 6-DMAP. Bead-bound proteins were analysed by immunoblotting. **c**, Nucleotide-depleted G2/M Glu extracts from YLD70 (see Methods) were incubated with peptide beads and the indicated additions (1 mM ATP,

3 mM 6-DMAP, 282 μM roscovitine, 10 mM EDTA) for 60 min at 30 °C. **d**, G2/M Glu extracts from YLD70 were incubated with peptide beads either in the absence (mock) or presence (+Cy) of 20 μg unbiotinylated Cy peptide, comprising amino acids 8–21 of CDC6(WT). **e**, Peptide binding was carried out as in **b** except that 6-DMAP was omitted. **f**, Extract from log phase HeLa cells was incubated with peptide beads as in **e**. **g**, Cy peptide was used in competition experiments as in **d**.

Cdc6 proteolysis in nocodazole-arrested cells¹². An integrating plasmid expressing Cdc6(WT) from its own promoter transformed both *ORC*⁺ and *orc2-6A, orc6-4A* strains with similar efficiency (Fig. 2d). Although Cdc6 Δ NT or Cdc6(1–4)A plasmids could efficiently transform the *ORC*⁺ strain, they were extremely inefficient at transforming the *orc2-6A, orc6-4A* strain, suggesting that the combination of the *orc* and *cdc6* mutants is lethal. We confirmed this by showing that haploid strains containing either Cdc6 Δ NT or Cdc6(1–4)A expressed from the *CDC6* promoter did not generate viable colonies when crossed to a haploid *orc2-6A, orc6-4A* strain (Fig. 2e). Cdc6(3,4)A was not lethal in combination with *orc2-6A, orc6-4A*, indicating that CDK sites 1 and 2 are sufficient for regulation through the Cdc6 NTD. This extends previous work showing that the combination of stabilized Cdc6 Δ NT and unphosphorylatable Orc6 is lethal¹⁴. We presume that lethality is due to some re-replication.

We were interested in determining how Clb2/Cdc28 bound so tightly to the NTD of Cdc6. Peptides corresponding to the NTD in various phosphorylation states (NT, NTP, NT12P, NT34P, NTA, NTP Δ Cy and ScramP) were synthesized and biotinylated on their N termini (see Fig. 3a for phosphorylation patterns of these peptides). These biotinylated peptides were bound to streptavidin-coated magnetic beads, incubated in a G2/M extract, and bound proteins were analysed by immunoblotting. Figure 3b shows that Cdc4

bound specifically to the fully phosphorylated NTP peptide as well as NT34P, consistent with previous two-hybrid analysis indicating that CDK sites 3 and 4 are important for Cdc4 interaction⁶.

Surprisingly, neither Clb2 nor Cdc28 bound to the unphosphorylated NT peptide; however, both proteins bound very efficiently to the phosphorylated peptides NTP, NT12P and NT34P. Neither Cdc4 nor Clb2/Cdc28 bound to ScramP, a peptide containing a randomized NTD sequence with four phospho-Ser/Thr residues, revealing that the phospho-Ser/Thr residues are not sufficient for binding. Figure 3c shows that Clb2, Cdc28 and Cdc4 can bind to the unphosphorylated NT peptide if the peptide is incubated in extract with ATP. This can be blocked with CDK inhibitors 6-DMAP and roscovitine, indicating that CDKs can interact with and phosphorylate the NT peptide. Cdc4 and Clb2 can then bind tightly to the phosphorylated peptide. Figure 3d shows that Clb2 and Cdc28 bound equally well to the NTP Δ Cy peptide, which lacks two Cy (also known as RXL) motifs implicated in cyclin interactions^{15,16}, and binding of Clb2/Cdc28 to NTP was not affected by the addition of an excess of a peptide containing the Cy motifs from Cdc6. Thus, the binding of CDKs to the NTD requires its phosphorylation but does not require the Cy motifs. Figure 3e shows that Clb3 can also bind to the phospho-NTD peptides.

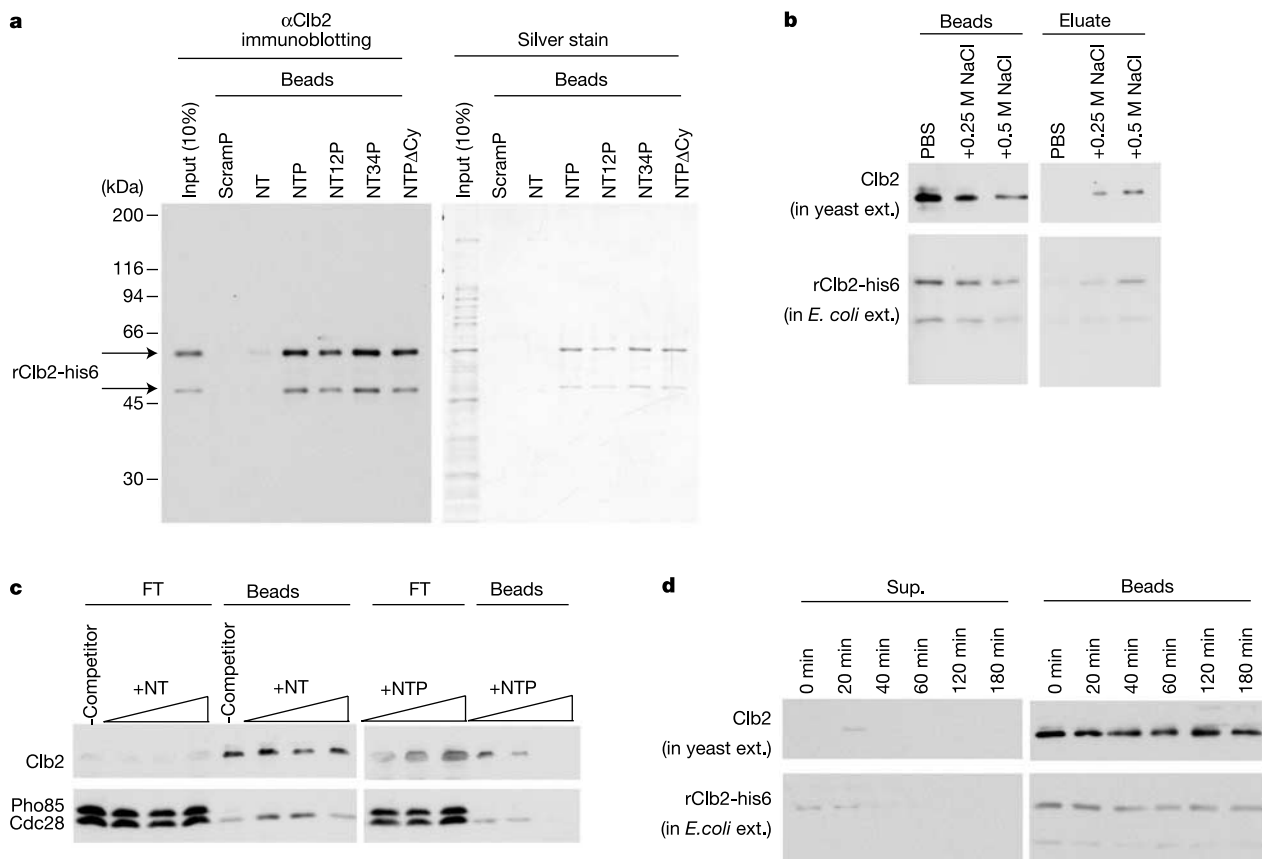


Figure 4 Recombinant Clb2 binds specifically to Cdc6 phosphopeptides. **a**, Extract prepared from *E. coli* expressing recombinant Clb2 was incubated with peptide beads (24 °C, 20 min), and bound proteins were analysed by immunoblotting and silver staining as indicated. **b**, NTP beads were incubated in either the *E. coli* extract or yeast G2/M Glu extracts (YLD70) at 24 °C for 20 min. Beads were isolated and incubated in PBS containing 0, 0.25 or 0.5 M NaCl. Bound and unbound proteins were analysed by immunoblotting. **c**, Peptide beads were incubated in G2/M extracts (YLD70; 3 mM

6-DMAP) and 0, 4, 8, 20 μ g of unbiotinylated NT or NTP peptides (24 °C, 20 min). Bound and unbound proteins were analysed by immunoblotting. **d**, NTP peptide beads were incubated in either the *E. coli* or yeast G2/M extracts (YLD70) at 24 °C for 20 min. Beads were isolated and then incubated in the presence of 20 μ g of unbiotinylated NTP for the indicated times. Bound and unbound proteins were analysed by immunoblotting.

Human cyclins A, B1 and E and the catalytic subunits Cdk2 and Cdc2, but not Cdk4 or Cdk6, bound specifically and tightly to NTP, NT12P and NT34P in HeLa cell extracts (Fig. 3f), suggesting that phosphopeptide binding activity has been conserved in evolution. The binding of cyclin A/Cdk2 to NT, NTA and to a lesser extent NTP was affected by competition from an excess of the Cy peptide (Fig. 3g). Cyclin B/Cdc2 binding to NTP was unaffected by the Cy peptide. Therefore, cyclin B1/Cdc2 and, to a lesser extent, cyclin A/Cdk2 also have phospho-NTD binding activity.

Experiments described in Supplementary Fig. 2 indicated that neither Cdc28 nor the small Cks1 subunit (homologue of Suc1) could account for phospho-NTD binding of the CDK complex. To address the role of the cyclin subunit in binding, extracts from *Escherichia coli* cells expressing recombinant Clb2 (rClb2) were used in pull-down experiments. Figure 4a shows that rClb2 specifically bound to the phosphorylated peptides NTP and NTPΔCy, as well as NT12P and NT34P. The silver-stained gel shows that Clb2 could be efficiently purified from *E. coli* extracts with NTP beads, but not NT or ScramP beads (Fig. 4a; silver stain). rClb2 and Clb2/Cdc28 from yeast extracts both remained bound to peptide beads even after treatment with 0.5 M NaCl (Fig. 4b). Clb2/Cdc28 from yeast bound to unbiotinylated NTP, efficiently competing with binding to biotinylated NTP beads when these peptides were added together (Fig. 4c). However, addition of excess levels of unbiotinylated NTP peptide after cyclins had already bound to biotinylated NTP beads had no effect on binding for at least 3 h (Fig. 4d), indicating that the Clb2–NTP complex of both *E. coli* and yeast extracts dissociates slowly. Because of the similarities between rClb2 and yeast Clb2/Cdc28 binding to NTP, we propose that binding of CDKs to the phosphorylated NTD of Cdc6 is mediated by the cyclin subunit.

Phosphorylation of Cdc6 by the G1 cyclins Cln1 and Cln2 generates two binding sites for SCF^{CDC4}—one in the NTD and one near the middle of the primary sequence—both of which contribute to targeting Cdc6 for very rapid, ubiquitin-mediated proteolysis during late G1 and S phases¹². During G2 and M phases, Cdc6 becomes more stable¹² but, as shown here, is still inactive. Cdc6 forms an extremely stable complex with Clb2/Cdc28 via its NTD, which cannot be loaded at DNA replication origins and cannot load Mcm2-7. We suggest that stabilization of Cdc6 later in the cell cycle^{12,17} is due to competition between mitotic cyclins and Cdc4 for sites 3 and 4 in the N terminus. This may help to ensure that Cdc6 made in mitosis is not immediately degraded but instead is maintained in an inactive form, allowing pre-RC assembly to be coupled directly to Clb2 degradation and mitotic exit.

The phosphopeptide binding activity of Clb2 was unexpected. We suggest that it may have a function in recruiting Clb2 to potential targets, perhaps promoting processive phosphorylation of substrates with multiple CDK sites. It will, therefore, be interesting to identify other proteins that bind to cyclins by this mechanism. The use of phosphopeptide binding domains to target kinases is an emerging trend in cell signalling^{18–21} and it will be interesting to know whether different cyclins have different binding specificities. If so, this may provide a new approach for interfering with the substrate targeting of different CDKs *in vivo*. □

Methods

Extract preparation

Yeast whole-cell extracts were prepared from cells arrested in either G1 (with alpha factor) or G2/M (with nocodazole) as described³. For the experiment in Fig. 3c, endogenous nucleotides were removed from the extract by two rounds of precipitation with 50% ammonium sulphate, re-suspension and dialysis (20 mM HEPES-KOH (pH 7.8), 300 mM K-glutamate, 160 mM sorbitol, 2 mM MgOAc, 0.4 mM EDTA, 0.8 mM dithiothreitol and protease inhibitors (1 mM AEBSE, 2 μg ml⁻¹ aprotinin, 1 mM benzamide hydrochloride, 10 μg ml⁻¹ leupeptin, 1 mg ml⁻¹ pepstatin A))²². HeLa cell extract was

prepared as described²³ with modification; proteins were extracted by 300 mM of K-glutamate before centrifugation. *Escherichia coli* extract was prepared from FB810 cells transformed with pET15b CLB2 construct. Cells were lysed by sonication in KGI20 (5 mM β-mercaptoethanol, 20 mM imidazole, 40 mM HEPES-KOH (pH 7.8), 160 mM sorbitol, 0.1% w/v Triton-X100, 2 mM magnesium acetate, 0.25 mM AEBSE, 600 mM K-glutamate). The cell lysate was centrifuged at 48,700 g for 20 min at 4 °C. For gel filtration, 100 μl of yeast extract or recombinant Cdc6 purified from *E. coli*²² were loaded onto a Superdex 200 HR 10/30 column (Pharmacia) pre-equilibrated with GFB (50 mM HEPES-KOH (pH 7.8), 300 mM K-glutamate, 200 mM sorbitol, 5 mM MgOAc, 1 mM EDTA). *In vitro* pre-RC assembly was assayed as described previously³ except that we used magnetic beads coupled to ARS1-containing plasmid instead of oligonucleotide beads²². p13^{suc1} treatment was performed by incubating 100 μl of G2/M extracts with 50 μl of either p13^{suc1} beads or BSA beads (a gift from T. Hunt) for 30 min at 4 °C, repeated twice.

Yeast techniques

The re-replication assay was performed as described¹³ with slight modification. Cells were grown in minimal medium containing 2% sucrose and required amino acids to log phase and arrested in M phase by incubating in YP- raffinose containing 2 mM methionine for 2.5 h at 30 °C. Then nocodazole was added and the cells were incubated for a further 30 min (+Met, Noc). Finally galactose was added to 2%, and cells were collected every hour after galactose addition (+Gal, 1 h, 2 h, 3 h and 4 h) for analysis by flow cytometry. For the transformation assay, cells were transformed with pRS306-based plasmids carrying a *CDC6* gene under its own promoter (see Supplementary Table). After 3 days, the numbers of URA⁺ colonies were counted. For the diploid formation assay, *MATa* strains (*ade⁻ his⁺ trp⁻ ura⁻*) and *MATα* strains (*ade⁻ his⁺ leu⁻*) were cultured either on YP-glucose or on SD-glucose containing adenine, leucine and 4 mM methionine for 3 days at 30 °C.

Immunoprecipitation

Yeast extracts and four volumes of GFB* (GFB plus protease inhibitors and phosphatase inhibitors (0.1 mM NaVO₃, 20 mM β-glycerophosphate, 1 mM NaF)) were mixed and centrifuged at 279,000g for 10 min at 4 °C. The supernatant was isolated and Triton-X100 was added to a final concentration of 0.05% w/v. The supernatant was incubated for 1.5 h at 4 °C with 30 μl of protein G-agarose beads (Sigma), which were incubated with either 50 μl of PBS or antibody solution. Beads were washed with 500 μl of ice-cold GFB* two times and 500 μl of ice-cold PBS.

Peptide pull-down analysis

Four micrograms of biotinylated peptides were bound to 20 μl of Dynabeads M-280 streptavidin (Dyna) in 50 μl of T₁₀E₁N₁₀₀₀ (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 M NaCl) for 30 min at room temperature. Peptide beads were then incubated with 0.2 mg of either yeast, HeLa cell or *E. coli* extracts that had been diluted with four volumes of GFB* for 20 min at 24 °C. The beads were washed with 400 μl of ice-cold GFB* two times. For off-rate measurements, the washed beads were further incubated in GFB* containing 20 μg of peptides at 24 °C. For high salt wash, the washed beads were further incubated in 200 μl of either ice-cold PBS, PBS plus 0.25 M NaCl or PBS plus 0.5 M NaCl for 2.5 min on ice two times.

Immunoblotting

Orc6, Mcm2 and Cdc6 were detected as described previously³. Cdc28 was detected with anti-PSTAIR antibody (a gift from J. Gannon) at 1:10,000. Endogenous Clb2, recombinant Clb2, Clb3 and Cdc4 were detected with sc-9071, sc-6699, sc-7167 and sc-6714 (Santa Cruz) antibodies at 1:250, respectively. Human cyclin A, B, Cdc2 and Cdk2 were detected with E23, V152, A17 and AN4.3 antibodies (a gift from J. Gannon) at 2 μg ml⁻¹, and human cyclin E was detected with sc-247 antibody (Santa Cruz) at 1:100. Protein A–HRP (Pharmacia), anti-mouse IgG–HRP (Vector) or anti-goat IgG–HRP (Santa Cruz) were used as secondary antibodies at 1:5,000.

Received 8 July; accepted 10 September 2004; doi:10.1038/nature03024.

Published online 20 October 2004.

- Diffley, J. F. X. Regulation of early events in chromosome replication. *Curr. Biol.* **14**, 778–786 (2004).
- Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
- Seki, T. & Diffley, J. F. X. Stepwise assembly of initiation proteins at budding yeast replication origins *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 14115–14120 (2000).
- Arion, D., Meijer, L., Brizuela, L. & Beach, D. cdc2 is a component of the M phase-specific histone H1 kinase: evidence for identity with MPF. *Cell* **55**, 371–378 (1988).
- Dunphy, W. G., Brizuela, L., Beach, D. & Newport, J. The *Xenopus* cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. *Cell* **54**, 423–431 (1988).
- Drury, L. S., Perkins, G. & Diffley, J. F. X. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J.* **16**, 5966–5976 (1997).
- Elsasser, S., Lou, F., Wang, B., Campbell, J. L. & Jong, A. Interaction between yeast Cdc6 protein and B-type cyclin/Cdc28 kinases. *Mol. Biol. Cell* **7**, 1723–1735 (1996).
- Weinreich, M., Liang, C., Chen, H. H. & Stillman, B. Binding of cyclin-dependent kinases to ORC and Cdc6p regulates the chromosome replication cycle. *Proc. Natl Acad. Sci. USA* **98**, 11211–11217 (2001).
- Archambault, V. *et al.* Genetic and biochemical evaluation of the importance of Cdc6 in regulating mitotic exit. *Mol. Biol. Cell* **14**, 4592–4604 (2003).
- Piatti, S., Bohm, T., Cocker, J. H., Diffley, J. F. X. & Nasmyth, K. Activation of S-phase promoting

CDKs in late G1 defines a "point of no return" after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes Dev.* **10**, 1516–1531 (1996).

11. Desdouets, C. *et al.* Evidence for a Cdc6p-independent mitotic resetting event involving DNA polymerase α . *EMBO J.* **17**, 4139–4146 (1998).
12. Perkins, G., Drury, L. S. & Diffley, J. F. X. Separate SCF^{CDC4} recognition elements target Cdc6 for proteolysis in S phase and mitosis. *EMBO J.* **20**, 4836–4845 (2001).
13. Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**, 1068–1073 (2001).
14. Wilmes, G. M. *et al.* Interaction of the S-phase cyclin Clb5 with an 'RXL' docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes Dev.* **18**, 981–991 (2004).
15. Adams, P. D. *et al.* Retinoblastoma protein contains a C-terminal motif that targets it for phosphorylation by cyclin-cdk complexes. *Mol. Cell. Biol.* **19**, 1068–1080 (1999).
16. Chen, J., Saha, P., Kornbluth, S., Dynlacht, B. D. & Dutta, A. Cyclin-binding motifs are essential for the function of p21CIP1. *Mol. Cell. Biol.* **16**, 4673–4682 (1996).
17. Drury, L. S., Perkins, G. & Diffley, J. F. X. The cyclin dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr. Biol.* **10**, 231–240 (2000).
18. Pawson, T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **116**, 191–203 (2004).
19. Elia, A. E. *et al.* The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* **115**, 83–95 (2003).
20. Cheng, K. Y., Lowe, E. D., Sinclair, J., Nigg, E. A. & Johnson, L. N. The crystal structure of the human polo-like kinase-1 polo box domain and its phosphopeptide complex. *EMBO J.* **22**, 5757–5768 (2003).
21. Durocher, D. *et al.* The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* **6**, 1169–1182 (2000).
22. Halford, K. A. *Biochemical Analysis of Yeast Pre-Replicative Complex Assembly*. Thesis, Univ. College London (2003).
23. Li, J. J. & Kelly, T. J. Simian virus 40 DNA replication *in vitro*. *Proc. Natl Acad. Sci. USA* **81**, 6973–6977 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to members of our laboratory for discussions and for critical reading of the manuscript. We are also grateful to J. Gannon and T. Hunt for discussions as well as human CDK antibodies and p13^{suc1} beads. We thank L. Drury, K. Labib, J. Li, G. Perkins and S. Reed for yeast strains. We also thank N. O'Reilly and the Peptide Synthesis Facility at the London Research Institute. This work was supported by Cancer Research UK and the Human Frontier Science Program Organization. S.M. is supported by JSPS postdoctoral Fellowships for Research Abroad.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.E.X.D. (john.diffley@cancer.org.uk).

corrigendum

Cloning of adiponectin receptors that mediate antidiabetic metabolic effects

Toshimasa Yamauchi, Junji Kamon, Yusuke Ito, Atsushi Tsuchida, Takehiko Yokomizo, Shunbun Kita, Takuya Sugiyama, Makoto Miyagishi, Kazuo Hara, Masaki Tsunoda, Koji Murakami, Toshiaki Ohteki, Shoko Uchida, Sato Takekawa, Hironori Waki, Nelson H. Tsuno, Yoichi Shibata, Yasuo Terauchi, Philippe Froguel, Kazuyuki Tobe, Shigeo Koyasu, Kazunari Taira, Toshio Kitamura, Takao Shimizu, Ryozi Nagai & Takashi Kadowaki

Nature **423**, 762–769 (2003).

In this Letter, Fig. 1 is an illustration of the sorting procedure, rather than an original data set, which we did not explicitly describe. Because the *x*-axis was used for FACS analysis of both FITC- and PE-labelled cells, only the gated population was shown; these data were extracted from the analyses and inserted into a FACS profile. The same file of a single gated population (from Fig. 1c) was mistakenly misused for a part of the plot in Fig. 1b by inserting it into both Fig. 1b and Fig. 1c in the original figure. In order to clear up the confusion surrounding this figure, the original primary data are shown here in the Supplementary Information. The Supplementary Information also includes further details of our cell-sorting procedure, which were not provided in the published protocol. Although these corrections do not affect the conclusions of our paper, we apologize to readers who have been misled by these mistakes. □

Supplementary Information accompanies the corrigendum on www.nature.com/nature.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

Pennsylvania, New York,

Maryland/ Latin America/ NIH:

Michael McGillion

Midwest USA:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

e-mail: p.phelan@nature.com

o.wulffen@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@nature.jp.com

naturejobs

Making the match

Communication. It's one of those vague yet lofty skills that scientific recruiters say are important, but few define, and fewer still can present concrete examples. In a survey of 77 scientists in academia, biotech and pharma, all emphasized the importance of communication skills and provided examples of putting them into practice, both to get a foot in the door (see *Naturejobs* Recruiters & Industry, page 1128) and to manage an interview (see next month's *Nature Biotechnology*).

Whatever the setting, the all scientists surveyed by Grace Wong, founder and chief scientific officer of Massachusetts biotech company ActoKine Therapeutics, and founder and president of the biotech education organization Student Vision, emphasized the importance of matching their skills with an employer's needs. Juerg Meier, executive director of Novartis in Basel, Switzerland, says the big mistake in an interview is over-emphasizing your own objectives. To counter that, advises Thomas Kindt, director of the Division of Intramural Research at the US National Institutes of Health, you should show willingness to work towards an institution's mission.

Of course, knowing what an institution's mission is requires another key step — preparation. David Baltimore, president of the California Institute of Technology, says interviewees should be ready to discuss the full range of implications of both their own and the interviewer's work. Paul Kassner, an Amgen research scientist, advises giving a practice seminar to peers inside and outside your area of expertise. Answering potential questions from friends and colleagues helps you prepare for the real thing — but in order to anticipate those questions, you must research the company's history and the contributions of its scientific staff.

Following these steps will turn communication from a soft, abstract concept into a reality that can help you land the next job.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Fast track: charting the
course of your postdoc p1126

CAREER VIEW

Recruiters & Industry

Get your foot in the door

Postdoc Journal

PhD limitations

Movers

Julia King

p1128

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Are you on course for the career you want? Don't follow the crowd and lose your direction, warns Eugene Russo. Instead, map out your own postdoc path.

Raymond Clark never wanted to be a postdoctoral fellow. Now, as a leader in the US National Postdoctoral Association (NPA), he listens to postdocs' concerns to try to change what many believe to be a flawed system.

After receiving his PhD in physiology from Idaho State University, Clark envisaged teaching and conducting research at an undergraduate-only institution. He didn't want graduate students. He didn't want to worry about pursuing grants. As a doctoral student, he'd had lots of independence and ample teaching experience. He'd served on university committees, and advised undergraduate and master's students. By the time he finished his degree, he felt ready to step into a faculty role. "I did everything a faculty member did. I was ready to go," says Clark. "Then I went into a postdoc and became a nonentity, especially with the culture and the lab I went into."

Clark did two stints in different labs at the University of California, San Diego, for a total of five years. "The postdoc in the life sciences has become the de facto terminal degree," he says. "The writing was on the wall: do a postdoc or you don't get a job as a faculty member at a decent school at almost any level." Like many budding biomedical scientists in the United States, Clark felt that he'd worked as a postdoc without any real direction or time frame in mind.

In recent years, the NPA, the National Academy of Sciences, the National Institutes of Health (NIH) and other organizations have worked to secure better pay, benefits, training and rights for postdocs in the United States. Discussions have also focused on ways to define the postdoctoral experience better and to stick with a specific timetable. Meanwhile, many postdocs — particularly in the life sciences — follow the crowd, with the often-flawed assumption that they'll be guaranteed a job at the end of their fellowship.

DIFFERENT FIELDS, DIFFERENT OPPORTUNITIES

The scenario varies from one field to another. In physics and chemistry, where postdoctorates are typically shorter than in the life sciences, it is closely correlated with the economy: the number of postdoc stints rises when unemployment does. Even in the life sciences, the statistics tell a complicated story. Despite the recent uproar about endless life-science postdocs,



Which way to go? The traditional route isn't always the best.

trends seem to be reversing: the length of first postdoctoral appointments among life-scientists seems to be decreasing (see H. H. Garrison, S. A. Gerbi & P. W. Kincade *FASEB J.* **17**, 2169–2173; 2003).

Regardless, many scientists are still frustrated by directionless, low-paid postdoctorates that offer little training. Now a policy committee co-chair at the NPA, Clark believes his experience was typical: the head of his lab wasn't interested in training a young scientist — she wanted skilled labour for her research project.

Some principal investigators (PIs) understand that training, leading towards independence, should be the priority. "Postdocs should not be copies of their advisers," says Keith Yamamoto, vice-dean for research at the Medical School of the University of California, San Francisco.

"I find great difficulty in understanding a postdoc





Forethought: Philip Clifford recommends it for setting postdoc goals; Lille Tidwell (below) used it to arrange her career path.

L. FEDORKOVA

who will go to a lab and will work on a project that's specifically to do with a PI's [grant]," agrees James William Nelson, a professor of cellular physiology at Stanford University School of Medicine in California. "Where's the independence in that?" Nelson urges his own postdocs to "use and abuse" his laboratory, in working towards an independent project that is likely to impress their prospective employers.



DOWN TO BUSINESS

Establishing independence isn't easy. But postdocs who don't leave their fates entirely in the hands of their PI are often able to navigate a successful career path.

Lille Tidwell credits careful planning and forethought for what looks to be a promising future in technology transfer. After earning her doctorate in neurobiology and anatomy, Tidwell struggled with her next move. Doing a postdoctorate would garner more publication credits, so she started one at Georgetown University in Washington DC. But she was wary of a medical-centre career in which much of her salary would come from her own grants.

A career seminar on technology transfer piqued her interest and changed her direction. She earned an NIH Office of Technology Transfer individual research training award, to serve at the NIH in Rockville, Maryland, and is very positive about her new career path. She will spend up to five years as a postdoc. "But my career has taken on a much brighter future," says Tidwell. Her awareness of the postdoc's plight led her to create the Georgetown

"You want to get out earlier and you want credit for the ideas you contribute. A PI may want you to get out later and not take knowledge to another lab."

— Richard Freeman

University Postdoctoral Association, and she was later elected to the executive board of the NPA.

Hoping to get postdocs to focus on their goals and move through the system more efficiently, the Federation of American Societies for Experimental Biology (FASEB) two years ago drafted an 'individual development plan' (IDP). One of its creators, Philip Clifford, professor of anaesthesiology at the Medical College of Wisconsin in Milwaukee, says it was born of the recognition that too many postdoctoral fellows had little sense of what they needed to accomplish. "They were just sort of following the crowd, doing the same thing everybody else had done with the expectation that there was going to be some sort of a job at the end of the postdoc rainbow," says Clifford. The IDP is geared towards biomedical science, but is of use to postdocs in any field, he says.

Modelled on the sort of priority-setting done in industry, the plan suggests, for example, that postdocs "conduct a self assessment" that analyses current abilities and outlines long-term objectives. It suggests identifying career opportunities with one's mentor and establishing effective dates for the length of a post-doctoral appointment. If a postdoc has an interest in industry, for example, then he or she might take steps to experience that environment. IDP users are asked to review their objectives every few months. "It just helps to formalize the process, makes people take it more seriously," says Clark. The IDP also offers a step-by-step process for mentors to make them more effective and aware of their obligations.

Harvard economist Richard Freeman, who researches the science and engineering workforce for the non-profit National Bureau of Economic Research, calls the IDP a nice idea, but too "touchy-feely". He believes postdocs need to accept that their interests — new skills, new research areas, independence — are often in conflict with their mentor's. "You want to get out earlier, you want credit for the ideas you contribute. A PI may want you to get out later and not take this knowledge to another lab," he says.

Freeman would like to see a short course included in the PhD programme to teach career management and what the postdoc's rights are. He'd also favour drafting a contract that states explicitly that the postdoc is there to both work in the lab and be trained.

One potential model could come from the world of business, where students are taught to manage their career path, as well as to organize other people into helping further the student's own objectives. This may

seem aggressive in science, but that's what PIs end up doing, says Freeman.

Postdocs are taught to look for the people doing the best research, which can be misleading, according to Clark. Finding out how long previous postdocs have been in the lab, and what the PI's interests are — if he or she has industry connections, say — can be just as important.

"They're not taught the way we teach business students, to look not only for the best business environment but for the best human resources, the best support and the best teams," says Clark. Successful, quick postdoctorates are most likely when students take at least as much responsibility for their own training as the institutions and faculty do.

Eugene Russo is a freelance science writer in Takoma Park, Maryland.

GRADUATE JOURNAL

PhD limitations

When I decided to pursue a PhD, I thought the long road to this short acronym would pay off with respect (mainly from people outside academia), more job opportunities and better pay, in academia and industry.

But I've now realized there are no guarantees — especially in 'nontraditional' jobs outside research. Recently I was talking to someone who had just gained her PhD and was trying hard to find a job in pharmaceutical marketing. She has discovered that people with doctorates are not readily hired in that field, as they are considered too expensive and over-qualified. But one of my fellow students who left university after his master's degree has got the sort of job she was seeking, in a leading pharmaceutical company.

A PhD reflects specialization in a research area and therefore imposes certain limits on future job choices. Looking for a research position is usually a matter of availability and complementary research goals, but diving into non-research subjects can be difficult and sometimes requires additional skills. Thinking about this earlier, rather than later, in your career can save you time. I plan to use these conversations to see where I want to go and what skills I need to get there — before I complete my PhD. ■

Philipp Angerer is a third-year PhD student in biotechnology at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

Get a foot in the door

There is no magic trick for landing a good job in science. But if you produce good research or file useful patents, you can take a few steps to plant the seeds of opportunity.

Network. Introductions from friends, mentors and collaborators are especially valuable. When you come in contact with people who teach you and share ideas, do reciprocate. Show a sincere interest in their achievements and help them during difficult times.

Follow up. Collecting business cards is even more important than giving out yours. Maintain close contacts with mentors and friends, not just when you need a job. E-mail your CV directly to scientists, who will notice your attributes more than administrators. If you cannot meet the speakers from conferences, search for their e-mail addresses and ask for slides. Volunteer. To build new

skills and contacts, spend some time in a different lab. Volunteering at commercial conferences can get you a free pass to attend seminars and meet speakers. Organizing seminars and panel discussions for local societies is also a good way to meet scientists.

Be industrious. Even if you are based in academia, collaborate with industry-based scientists. They could provide unique reagents or helpful recommendations. One may even become your future boss. Consider a temporary industry job or a postdoc. You will gain valuable perspective on industry, and may get hired permanently.

Meet people. Attend free seminars at small academic symposia to meet scientists and catch up on hot technologies. Talk to vendors at trade shows. Ask them which companies are hiring and what technologies are hot. Use your first meeting to establish rapport: talk about science or shared

interests rather than jobs.

Prepare. Have a short, memorable 'smart pitch' ready to market your scientific background and accomplishments in simple terms a broad audience can understand, and in less than a minute. Also tailor your resumé or CV to match a company's needs or its job descriptions. If possible, send it directly to a scientist you've met in the company — it's much more likely to be read.

Above all, remember that all interactions are potentially important. Be cheerful, kind and helpful to everyone, not only managers! Show a passion for science. Plant many seeds, learn, become wiser today than yesterday and create more options. A positive-minded problem-solver with creative ideas, talents or expertise will be welcome at any door. ■

Grace Wong is founder and chief scientific officer of ActoKine Therapeutics and founder of Student Vision.

MOVERS Julia King, principal of the engineering faculty, Imperial College London



It was the champagne lifestyle that first attracted Julia King to a career in science. During the optimistic 1960s, says Imperial's new head of engineering, the newspapers seemed full of pictures of scientists celebrating their latest breakthrough with a bottle of bubbly.

"Particle physics was the exciting face of science when I was young," says King, who was born the year the

European particle-physics lab CERN was set up: both have just celebrated 50.

At the University of Cambridge, where she went to study physics, King discovered new interests, but the switch to materials was painless. For that reason, she says, "I'm a strong supporter of courses such as natural sciences at Cambridge, where you can sample a range of things".

Despite the good pay and prospects on offer, and the excitement of seeing one's ideas become a product, King can see reasons why women haven't made more inroads into engineering. These include loneliness, the work-life balance — she and her husband are enjoying living together after years of weekend commuting — and a lack of role models that could make a woman's ambition seem somehow inappropriate.

"Engineering can still be quite an oppressive environment for a woman, though not intentionally," she explains.

"When you're finding it tough there isn't an obvious person to have a chat with, and you may not see that, in other parts of the company, women have made it up the management ladder."

Excellent female science teachers were her own first role models, but she advises students to be open to guidance from all sources. One of the most useful pieces of advice she received was from an old technician at a US aeroengine maintenance plant, who said the engine dressings — an intricately designed mass of pipes and wires — gave his team hours of extra work when they had to be dismantled. "That taught me to think of what's useful to the customer," she says.

She's keen on an area that could draw into engineering some of the young female science students who currently head for medical school. At Imperial's new bioengineering department, 50% of undergraduates are women. And as she says, "That's something to celebrate". ■

CV **2002–04:** chief executive, Institute of Physics
1994–2002: Rolls-Royce plc: head of aerospace materials; managing director, fan systems; director of marine engineering & technology
1992–94: lecturer, University of Cambridge
1987–92: British Gas/Royal Academy of Engineering senior research fellow, University of Cambridge
1980–87: lecturer, University of Nottingham
1978–80: Rolls-Royce research fellow, Girton College, Cambridge
1972–78: BA in natural sciences and PhD in fatigue and fracture, New Hall, Cambridge